Improvement of methodologies for experimental charge density determination

Natalie Johnson

Submitted for the qualification of Doctor of Philosophy

School of Chemistry

May 2019

Abstract

For over 100 years, crystallography has utilised the interaction of X-radiation and electrons within a crystal to enable the reconstruction of a 3-dimensional model of the arrangement of atoms within crystalline solids. Crystallography is considered to be the gold standard of analytical techniques within the chemical and material fields.

An atomic model based on the distribution of the electron density is most commonly generated using a number of parameters to describe the atomic arrangement within space. A standard crystallographic experiment assumes that electrons are spherically distributed around atomic nuclei. However, this does not account for electron density within bonding regions, or other classically known features such as lone pairs of electrons. Charge density studies, the pinnacle of structural determinations, involve the addition of parameters to a structural model to characterise the non-spherical density. This can improve the accuracy of the model of the overall distribution of the electron density and allow further investigation into bonding, intermolecular interactions and more advanced structural properties.

Some studies into the reproducibility of charge density models have previously been attempted however these are far from exhaustive. The reproducibility, i.e. how well features in the models can be recreated when using different instrumentation and data processing methods, infers the dependability of the parameters derived from models and is therefore paramount. Many different options for processing data exist, including the choice of integration algorithm and if any diffraction data should be excluded from a refinement. These choices may affect the final form of the model of electron density distribution and hence any properties derived from it and therefore need to be approached with caution. Within this thesis an investigation into the effect of processing methods is undertaken using 2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane.

Additionally, newly developed and improved methods to assist the processing and evaluation of charge density are discussed in detail including a program for calculating optimised parameters for a weighting scheme that can be applied to charge density refinements. A number of small molecule single crystal structures have been investigated and their structural models are presented. These highlight the use of the newly developed tools and methodologies along with their effect on high resolution charge density studies.

To all those who believed I could do it

Acknowledgements

I would first like to thank my supervisor Dr Michael Probert for providing me with the opportunity to work in a challenging area of crystallography. I am grateful for L-ascorbic acid data provided by Dr Holger Ott at Bruker AXS GmbH and to Dr Jonathan Sellars for providing the sample of DCMC. Thanks also to Dr Mark Warren at Diamond Light Source for presenting me with the challenge of converting diffraction data frames, the assistance of Dr Markus Gerstel for all things relating to the operation of the Diamond Light Source computer cluster and to Dr Dmitry S Yufit for providing helpful feedback on the *cbf_to_sfrm* software. Additional thanks to Professor Bill Clegg for his helpful explanations of crystallographic techniques throughout my time at Newcastle.

To the Newcastle University Irish Dancing Society, thank you all for keeping me sane and giving me the opportunity to perform all over North East. I had a reel-y great time dancing with you guys.

To my friends, especially Caitlin, Christina, Kristen, Leanne, Saranja and Shelby who have patiently listened and offered advice over all the trials of the past four years, thank you. I am incredibly lucky to have you all in my life.

Thanks to all the members of Data Handling Area team for their support, especially Jake Musselle-Sexton. The most heartfelt thanks go to Dr Paul Waddell for his unwavering moral support and assistance. Thank you for listening to my problems and keeping me entertained over the course of my studies. The real lab-lab was inside us all along.

And finally, to my parents, my grandmother, and my brother, thank you for all your support and encouragement over the last four years. I am forever grateful.

Contents

Abbreviations

ADP	Atomic displacement parameter
BCP	Bond critical point
CCD	Charge coupled device
cctbx	Computational crystallographic toolbox
CIF	Crystallographic information format
CMOS	Complementary metal-oxide semiconductor
CPAD	Charge integrating pixel array detector
CSD	Cambridge Structural Database
D	Dimensions
DCM	Dichloromethane
DCMC	1,4-bis(dicyanomethylene)cyclohexane
DMSDA	Differences of the mean square displacement amplitude
F_2TCNQ	2,5-difluoro-7,7,8,8-tetracyanoquinodimethane
F_4TCNQ	2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane
GooF	Goodness of fit
GUI	Graphical user interface
HPAD	Hybrid pixel array detector
IAM	Independent atom model
IUCr	International Union of Crystallography
SHADE	Simple hydrogen anisotropic displacement estimator
TAAM	Transferable aspherical atom model
TCNQ	7,7,8,8-tetracyanoquinodimethane
TDS	Thermal diffuse scattering
TTF	Tetrathiafulvalene
UBDB	University at Buffalo databank
wGooF	Weighted goodness of fit

Chapter 1. Introduction

X-ray crystallography is a powerful analytical technique that enables the investigation of the 3-dimensional (3D) atomic structure of a sample. The atomic structure of a sample is extremely important in the establishment of structure-property relationships.[1] X-ray crystallography can be used to probe a range of different compounds from small molecules and inorganic structures to biologically active molecules and large proteins.[2–4] This thesis focuses on so called 'small molecule crystallography', specifically single crystal studies with small organic molecules in the range of up to 25 atoms in the molecule.

1.1 Basic theory for crystallography

A single crystal has an ordered 3D structure and is made of repeating units, like building blocks, that are identical to each other. The smallest repeated building block that can be used to construct the whole crystal, through only translation of the block, is called the unit cell. Therefore, due to the repeating internal structure of the crystal, the entire structure can be described just by using this unit. The way molecules are arranged in space and the symmetry that arises as a consequence of that arrangement can be described by the space group of the crystal – one of 230 different combinations of symmetry elements that a structure may possess, for systems that exhibit a 3D lattice. Additional symmetry within the unit cell may allow a smaller unique section of the structure to represent the entire crystal; the rest of the unit cell is then generated through applying the symmetry operations. The unique section is called the asymmetric unit.

The kinematic theory of diffraction states that when a beam of X-rays is directed onto a crystal the electric fields created by electrons within a crystal cause the beam to be scattered.[5] This can occur elastically, with no loss of energy (Thomson scattering), or inelastically, with a small loss of energy due to a transfer of energy to an electron (Compton scattering).[6] The emitted waves from all the electrons within the crystal interfere both constructively and destructively with each other to produce a diffraction pattern. The diffraction pattern of a well-ordered single crystal occurs as a series of discrete spots, which

can then be interrogated to produce an accurate representation of the atomic structure of the single crystal.

Lawrence Bragg proposed an analogy that diffracted beams could be modelled as if they were "*reflections*" from a group of imaginary parallel planes within the crystal.[7,8] These sets of planes can be described by their Miller indices (*h, k, l*), which are the reciprocal of the fractional coordinates at which the plane intercepts the unit cell axes. A pictorial representation of Bragg's law in 1D is shown in Figure 1.1.

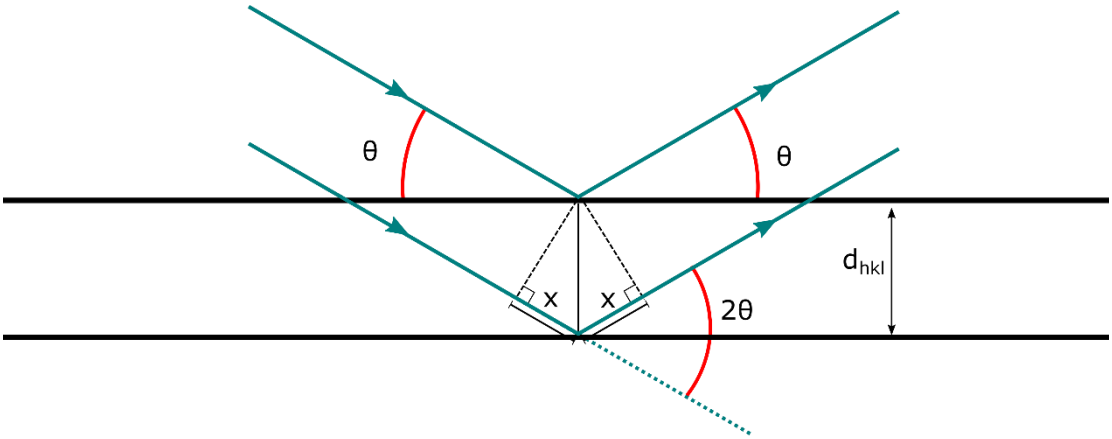


Figure 1.1 A pictorial representation of Bragg's law.

In order for constructive interference to occur, the reflected beams must be in phase. For this to happen the path difference between the beams, twice the distance labelled x in Figure 1.1, must be an integer multiple of the wavelength of the incident radiation as described by the Bragg equation (Equation 1.1), where n is number of wavelengths, λ is wavelength of X-ray beam, θ is the angle at which the beam hits the plane and finally d_{hkl} is the distance between adjacent planes.

$$n\lambda = 2d_{hkl}\, sin\, \theta \qquad 1.1$$

Each reflection can be represented by a vector of magnitude $1/d_{hkl}$ and a direction normal to the reflection planes. The vectors from different sets of planes can be combined together to form the reciprocal lattice. The reciprocal lattice contains information about the geometry of the crystal in 3D and can be calculated directly from an experimental diffraction pattern by calculating non-coplanar vectors between reflections.

The aim of most crystallographic experiments is to produce a model of the system, usually through the creation of a representative electron density map. The electron density of the crystal, p, at a point (*x, y, z*) in the unit cell is the sum the structure factor, F_{hkl}, for each reflection and an exponential term, which will be explained later (Equation 1.2).

$$p(xyz) = \frac{1}{V}\sum_{hkl} F_{hkl} e^{-2\pi i(hx+ky+lz)} \qquad 1.2$$

The structure factor equation (Equation 1.3) defines the intensity of a reflection as the sum of the waves scattered by every individual atom in the structure, with positions (x_j, y_j, z_j), multiplied by the function of their atomic scattering factors, f_j, the value of which depends on the direction of the scattered reflection. The diffraction pattern of a sample and the electron density of a crystal are related to each other by a Fourier transform.

$$F_{hkl} = \sum_{j=1}^{N} f_j \, e^{2\pi i(hx_j+ky_j+lz_j)} \qquad 1.3$$

Crystal structure analysis can be split into 3 stages: data collection, data reduction and structure solution. Modern single crystal diffraction experiments rely heavily on automated data collection routines and semi-automated data processing built into the instrument control software. This is a seemingly welcome aspect for most users, reducing the high level of knowledge for the processing of experimental data, as default settings of the software often produce sensible results for spherical atom models.

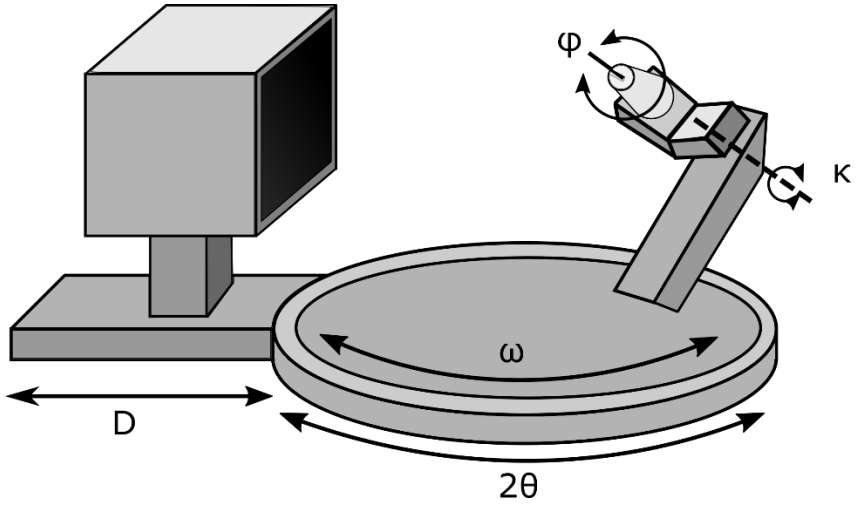


Figure 1.2 Diagram of a goniometer setup where D is the distance of the detector from the sample and θ, κ, φ and ω are the setting angles for a goniometer in Kappa geometry.

The diffraction pattern produced by a single crystal is measured during the data collection stage using a diffractometer. To enable this, a single crystal is mounted on a goniometer which can rotate the crystal around a variety of axes, as shown in Figure 1.2. The angles present on the diagram (θ, κ, ϕ and ω along with the distance of the detector from the sample, D) are the setting angles. There are a number of different geometries of goniometers available in modern crystallography including Eulerian, fixed chi and kappa. The goniometer shown in the Figure 1.2 has a kappa geometry.

When a beam of X-rays is directed at a crystal, the beam is scattered in many directions. The interference pattern formed by individual scattered beams causes the diffraction pattern to occur. In a standard diffraction experiment, the aim is to measure the intensity of reflections to a given resolution and redundancy. This requires the diffraction pattern to be recorded at a variety of setting angles. Diffraction patterns are symmetrical; therefore, the whole sphere of the diffraction is not needed to achieve these targets.

A crystallographic experiment normally requires multiple experimental 'runs' to collect enough information to be used to determine the crystal structure. For the newest instrumentation, an experimental run can be defined as the collection of data during the continuous rotation of one of the diffractometer axes (usually ϕ or ω). During an experimental run, the detector measures the diffraction pattern throughout the rotation of the chosen axis. The output is most often a diffraction frame, a file which contains an image showing the intensity of the resultant X-rays for a small fraction of the rotation of the axis; for small molecule crystallography, this is usually no more than a few degrees and often much smaller. Therefore, one experimental run is made up of many frames. Alongside the image, the frame also contains significant amounts of meta data for the experiment – such as the setting angles, time length and temperature at which the frame was collected.

To begin processing the diffraction data, an initial unit cell is calculated by identifying the three shortest non-coplanar vectors between reflections that allow the assignment of integer values for the Miller indices for all reflections; this unit cell is the initial reduced cell.[9–11] A number of other factors such as the metric and Laue symmetry and the presence of systematic absences throughout the dataset may affect the final unit cell parameters. The unit cell is described by six parameters, three lengths (a, b, c) and three angles (α, β, γ) which form a parallelepiped.

The intensity of the reflections in the diffraction pattern, and corresponding standard uncertainty values of the measurements, can then be extracted from the collected diffraction data during the integration stage. A series of computational algorithms are used to calculate the intensities of each reflection through the summations of pixel intensities around the calculated position of a reflection.[9,12]

Direct determination of the electron density from experimental data requires both the intensity of the reflection and its phase. Equation 1.4, an alternative representation of

Equation 1.3, shows that the structure factor of a reflection, F_{hkl}, is a combination of the intensity value and an exponential term containing the phase information. However, in a diffraction experiment, only the intensity of the reflection can be measured; the phases of the reflections are lost, giving rise to the 'phase problem'. Phases must therefore be estimated or recovered in order to allow the determination of the distribution of electron density from the experiment. This can be achieved using implicit relationships that are encoded into the structural model; relationships between reflections and their phases allow the estimation of phases for an approximate model of the electron density,[13] which is then iteratively improved upon during the refinement of the structure solution. A model of the electron density distribution is required as we cannot directly recreate the electron density distribution of the unit cell of a crystal from the diffraction pattern.

$$F(hkl) = |F(hkl)| \cdot e^{(i\phi(hkl))} \qquad 1.4$$

Structure factors are calculated for the model, F_c, and are compared to the observed structure factors (using the calculated phases), F_o. Least squares refinement of a model aims to minimise the squared difference between observed and calculated structure factors. Provided that the data recorded are of a high enough quality and resolution, the final outcome should be a well resolved model, showing the positions of all the atoms and therefore the molecular structure within the unit cell. Due to the timescale of a crystallography experiment in relation to atomic motion, the model created is that of the time averaged structure of the crystal. Additionally, the result is space averaged as the unit cell may have minor differences throughout the crystal.

1.1.1 Spherical atom model

A structural model is used as a way to describe the distribution of electron density within the unit cell. In the spherical or independent atom model (IAM), the electron density is assumed to be spherically located around the nucleus of the atom. This model is routinely used to provide a representation of the electron density and hence the crystal structure by approximating the electron density as functions centred at atomic positions within the unit cell.

A number of parameters are required to describe the density. These are the relative position of the centroid of the atom (*x, y, z*) and a description of the displacement of those positions either due to internal vibrations within the molecule (exacerbated by temperature) or misalignments of the atoms within the crystal, which are also known as anisotropic

displacement parameters (ADP)s. Up to six parameters are used to describe an ellipsoid of probability about the centre of the atomic nucleus and create the U_{ij} matrix. The ADPs can be either isotropic (U_{ii}, one parameter) or anisotropic (U_{ij}, 6 parameters).

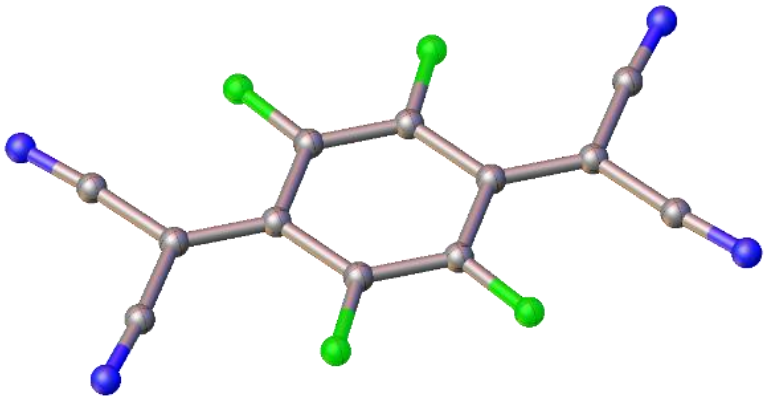


Figure 1.3 Model of a molecule showing bonds as drawn by a structural refinement program (Olex2).[14]

The bonds displayed in a spherical atom depiction of a crystal structure (Figure 1.3) are assigned based on geometric considerations and inter-atomic distances, along with expected coordination. All atom-atom distances that are less than expected bond lengths according to an arbitrary definition will be displayed as such – in most refinement software a user can fix the maximum connectivity for each atom or remove connections which look spurious.

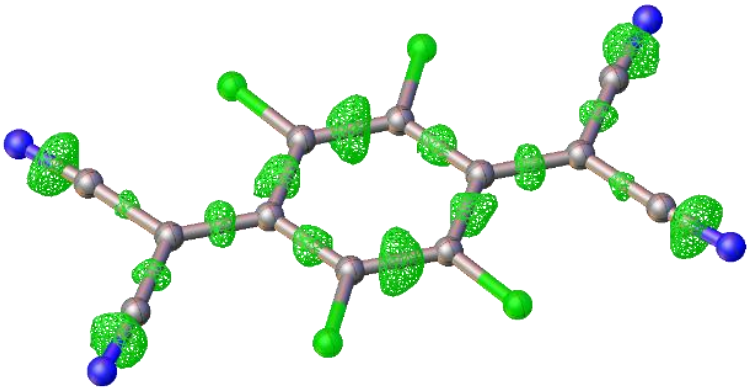


Figure 1.4 A molecule with unmodeled density calculated at 0.1 eA^{-3} level (Olex2).[14]

In a covalent bond it is expected that electrons will be shared between the atoms in order to form a chemical bond. When higher resolution data are available it may be possible to see evidence of this bonding within the unmodeled or residual electron density. This is calculated as the density arising from the difference between the observed and calculated structure factors. Figure 1.4 shows the unmodeled residual electron density for a molecule for which data were measured to 0.48 Å resolution; the surface is of the residual electron density calculated at 0.1 eÅ^{-3} level is shown as a green mesh. Residual electron density at this level is present between many of the atoms where covalent bonds may be expected as

indicated by a 3D molecule viewing program, *Olex2*.[14] Lone pair density on certain atoms can also be identified in this way.

The presence of residual electron density in potentially chemically meaningful locations indicates there is additional information in the diffraction pattern which is not described by the IAM. Modelling the bonding and lone pair electron density can provide useful insights in terms of drug design[15] and further the understanding of the structural properties of molecules.[1,16,17] Moreover, the creation of a more representative model of the distribution of electron density within a molecule (and a unit cell) allows quantitative analysis of molecular properties through the calculations of dipole moments, molecular charges and electrostatic potentials and the strength of electrostatic interactions from experimental diffraction data.[1] One method of modelling non-spherical density in the unit cell is the multipole model, or so called 'charge density' analysis, approach. This method underpins the work expressed in this thesis.

1.1.2 Multipole model

The multipole model, also known as the aspherical atom model, is a more complex model than the IAM. It is often known as multipolar expansion, as the spherical atom model has been expanded to include additional parameters to define the aspherical electron density, including density which is not located at the atomic site. As with the spherical atom model, these additional parameters are centred on the positions of atomic nuclei using spherical harmonic functions.

One of the most widely used formalisms was developed by Hansen and Coppens.[18] Electron density is divided into three parts: spherical core electron density, spherical valence density and aspherical valence density respectively (Equation 1.5).

$$\rho(r) = P_c\rho_c(r) + P_v\kappa^3\rho_v(\kappa,\mathrm{r}) + \sum_{\iota=0}^{l_{max}} \kappa'^3 R_l(\kappa',r) \sum_{m=0/-l}^{+\iota} P_{lm\pm} d_{lm\pm}(\vartheta,\varphi) \qquad 1.5$$

The population, *P*, of each term depends on the number of core and valence electrons within each atom, where P_c is the number of core electrons and the sum of P_v and P_{lm} is the number of valence electrons. The valence terms have expansion and contraction parameters, κ or κ', which can be refined to improve the fit of the functions to the density. The aspherical valence density has been modelled using many coefficients, including R_l, Slater-type functions and d_{lm}, density-normalised spherical harmonics. The form of the

multipolar functions chosen to describe the aspherical valence density mimic those of atomic orbitals from linear combination of atomic orbital theory.[19]

1.1.3 Quantum theory of atoms in molecules

The topology of the multipole model can be analysed to derive properties from the distribution of electron density within the model.[20] Points on the surface of the electron density where the gradient of a tangent to the density at that point is zero (Equation 1.6) are very important to the analysis of the electron density; these are known as a critical points and are used heavily to describe features within the distribution.

$$\nabla\rho(r) = i\frac{\partial\rho}{\partial x} + j\frac{\partial\rho}{\partial y} + k\frac{\partial\rho}{\partial z} = 0 \qquad 1.6$$

A representation of critical points for a 2D grid of values can be shown in 3D. In Figure 1.5, the graphs show a local maximum, a local minimum and a saddle point (where the critical point is at minimum in one direction, but a maximum in the other).

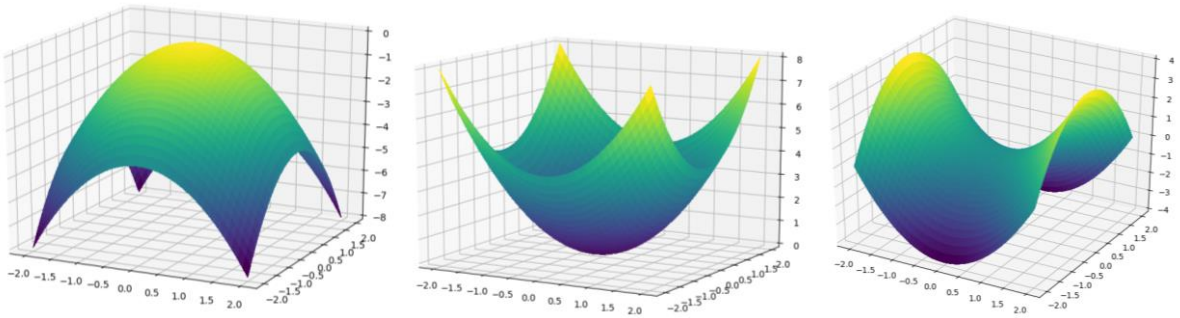


Figure 1.5 Critical points in 3D (where third dimension is the intensity): local maximum, local minimum and a saddle point.

To determine the form of the critical point (whether maximum, minimum or saddle point) the second derivative of the density must be calculated. This gives the curvature of the critical point and is determined using the 3 x 3 Hessian matrix, a matrix of partial second derivatives of the electron density, shown in Equation 1.7.

$$\mathrm{H(r)} = \begin{pmatrix} \frac{\partial^2\rho}{\partial x^2} & \frac{\partial^2\rho}{\partial x\partial y} & \frac{\partial^2\rho}{\partial x\partial z} \\ \frac{\partial^2\rho}{\partial y\partial x} & \frac{\partial^2\rho}{\partial y^2} & \frac{\partial^2\rho}{\partial y\partial z} \\ \frac{\partial^2\rho}{\partial z\partial x} & \frac{\partial^2\rho}{\partial z\partial y} & \frac{\partial^2\rho}{\partial z^2} \end{pmatrix} \qquad 1.7$$

The Hessian matrix can be diagonalized to give the Laplacian of the density (Equation 1.8).

$$\nabla^2\rho = \frac{\partial^2\rho}{\partial x^2} + \frac{\partial^2\rho}{\partial y^2} + \frac{\partial^2\rho}{\partial z^2} \qquad 1.8$$

Critical points can then be described by their rank and signature (ω, σ). The rank is the number of non-zero eigenvalues in the Hessian matrix. The signature is the sum of

eigenvalues. There are four possible combinations of signature for a rank of 3, which is expected for stable molecules.

- (3, -3), where all curvatures are negative. This is a local maximum, which is expected at atomic positions.
- (3, -1), a bond critical point (BCP), with two negative curvatures and one positive curvature. These are expected to be present between any two atoms that share a covalent bond, although the presence of a BCP between two atoms does not mean that there is a bond between the two atoms.[21]
- (3, +1), ring critical point, two positive curvatures and one negative curvature.
- (3, +3), cage critical point, all curvatures are positive, and therefore the critical point is a local minimum.

The Laplacian at the BCP provides information on the local concentration or depletion of charge density at that point. The sign of the Laplacian can also be used for determining whether a BCP relates to the formation of a covalent bond or indicates an ionic or van der Waals interaction.

- $\nabla^2\rho < 0$ signifies a charge concentration – this occurs in a covalent bond (also called an open shell interaction) where sharing of electrons between two atoms results in an increase of charge at the BCP.[1]
- $\nabla^2\rho > 0$ indicates a charge depletion at the BCP signifying an ionic or van der Waals interaction (a closed shell interaction), as the electron density is not shared between the atoms in these interactions and instead contracts towards the nuclei.[21]

Another property which can be calculated at BCP is the ellipticity, ε . This value is calculated using Equation 1.9, where λ_1 and λ_2 are the two eigenvalues from the Hessian matrix which correspond to the directions perpendicular to that of the bond.

$$\varepsilon = \frac{\lambda_2}{\lambda_1} - 1 \qquad 1.9$$

The ellipticity is sometimes used as a measure of 'π' bond character, although this does not hold for triple bonds. It is actually a measure of the cross-sectional shape of the bond – single and triple bonds have cylindrical distributions of electron density giving a ratio of eigenvalues close to one and an ellipticity value of 0; while for double bonds, where the electron density is distributed elliptically, the ratio of eigenvalues becomes > 1 resulting a non-zero ellipticity value. Ellipticity can be used to demonstrate the amount of conjugation

within a system as the conjugation of bonds causes single and triple bonds to deviate from the approximately spherical distribution of electron density at the BCP.[22]

Expected values of ellipticity were calculated by Bader *et al.*[23] using STO-3G level of approximation and are often used as reference values within charge density studies (Table 1.1). Both single and triple bonds as found in propane and methylacetylene are calculated as having zero ellipticity at the BCP. By contrast, the ellipticity at the BCP of the double bond in propene is calculated as 0.84. Butadiene is a conjugated molecule, with alternating double and single bonds, therefore the single bond has an increased ellipticity which is explained as the gaining of 'double bond character' due to conjugation – this also results in a reduction of the ellipticity at the BCP of the butadiene double bond when compared to the (almost) isolated double bond in propene. In benzene, conjugation occurs in all bonds, thus a reduced ellipticity when compared to an (almost) isolated double bond is reported at all BCPs.

Molecule	Bond	ε
propane	CH_3-CH_2CH_3	0.0
methylacetylene	$CH_3C{\equiv}CH$	0.0
propene	$CH_3CH{=}CH_2$	0.84
benzene	C_6H_6	0.34
1,3-butadiene	CH_2CH-$CHCH_2$	0.10
1,3-butadiene	$CH_2{=}CHCHCH_2$	0.72

Table 1.1 Table containing calculated ellipticity for a series of bonds. Adapted from Bader et al.[23] Copyright (2018) American Chemical Society.

1.2 Validation of charge density models

There are several measures that can be applied to ensure that completed charge density models are reasonable and to assess the quality of data that has been used to produce them. Some of the measures described below are used across all crystallographic data collections and refinements whereas others are specific to charge density datasets. The methods of validation can be separated into four categories: descriptors of internal consistency of data, model fit, error distributions in the multipole model, and model reasonableness. As stated by Henn and Meindl[24] – for a measure to be a quality indicator of the data there must be an associated reference value, either implicitly given or assumed. Therefore, all the measures described below will be deemed data descriptors unless otherwise stated.

It is also important to understand the various sources of error that crystallographic data and refinements are prone to: random and systematic error. Random errors are ones that are

not correlated to any property of the reflection (such as intensity, resolution etc.) and are often typified by their Gaussian distribution, although intensity measurements actually have a Lorentzian distribution due to counting statistics.[25] Repeated measurements of intensities can improve the precision of the value, how reproducible a measurement is, minimising the effect of these random errors.[26] Systematic errors, however, are correlated to properties of the reflection and taking repeated measurements of a value will not address this type of error. The accuracy of a measurement, how close the value measured is to the 'true value', only improves with the elimination of errors from the experiment, therefore, multiple measurements may increase the confidence in a poor measurement in this case if the errors are not taken into account. Failure to correct for errors introduced in the data collection, or the inclusion of the error during the data processing stage, impact the refinement and ultimately the final model and all parameters that are derived from it.

1.2.1 Descriptors of internal consistency

In crystallographic data, internal consistency descriptors express the degree of agreement between multiple measurements of the same reflection and/or its symmetry equivalent. The R_{int} value (Equation 1.10) is a measure of internal consistency of the intensity measurements i.e. how well repeated measurements of the same reflection agree with the average value of intensity.

$$R_{int} = \frac{\sum_{hkl} \sum_i |F_o^2 - \langle F_o^2 \rangle|}{\sum_{hkl} \sum_i F_o^2} \qquad 1.10$$

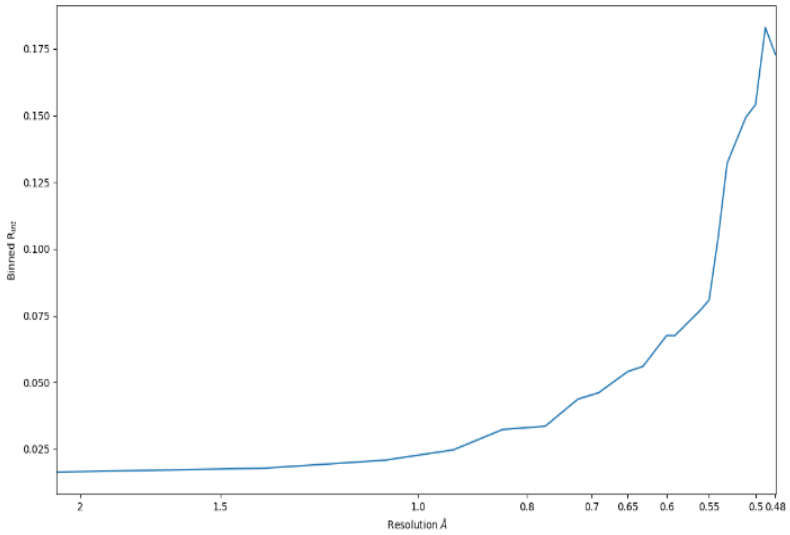


Figure 1.6 A graph of binned R_{int} with respect to the resolution of the reflection.

A graph of R_{int} vs resolution (Figure 1.6) has been suggested as a descriptor of the quality of collected data that provides insight into any resolution based trends in R_{int}.[27] This graph is

routinely produced during scaling stage of refinement by the absorption correction program *SADABS.*[28] Wolf *et al.* noted that data unsuitable for charge density refinement in their study of four datasets of [2,2]-paracyclophane could be identified as having R_{int} of higher than 5% for innermost resolution shells (∞ to 1.1 Å).

R_{int} is known to increase with increased multiplicity of measurements if the distribution of the measured values for each reflection is Gaussian,[29] therefore two new descriptors, $R_{p.i.m.}$ and $R_{r.i.m.}$ (Equations 1.11 and 1.12), were developed and take the multiplicity of a reflection, N, into account.[30,31] The $R_{p.i.m.}$ (precision indicating merging) factor was designed to describe the average precision of the measurement of the reflections, while the $R_{r.i.m.}$ (redundancy independent merging) factor gives the individual precision of each measurement. $R_{p.i.m.}$ and $R_{r.i.m.}$ as a function of resolution are also quoted as a measure of data quality at the scaling stage. The inclusion of a term to represent the redundancy should give a smaller value than R_{int} if the distribution of multiple measurements of the same reflection is Gaussian.

$$R_{r.i.m.} = \frac{\sum_{hkl}\left(\frac{N}{N-1}\right)^{\frac{1}{2}}\sum_i |F_o^2 - \langle F_o^2\rangle|}{\sum_{hkl}\sum_i F_o^2} \qquad 1.11$$

$$R_{p.i.m.} = \frac{\sum_{hkl}\left(\frac{1}{N-1}\right)^{\frac{1}{2}}\sum_i |F_o^2 - \langle F_o^2\rangle|}{\sum_{hkl}\sum_i F_o^2} \qquad 1.12$$

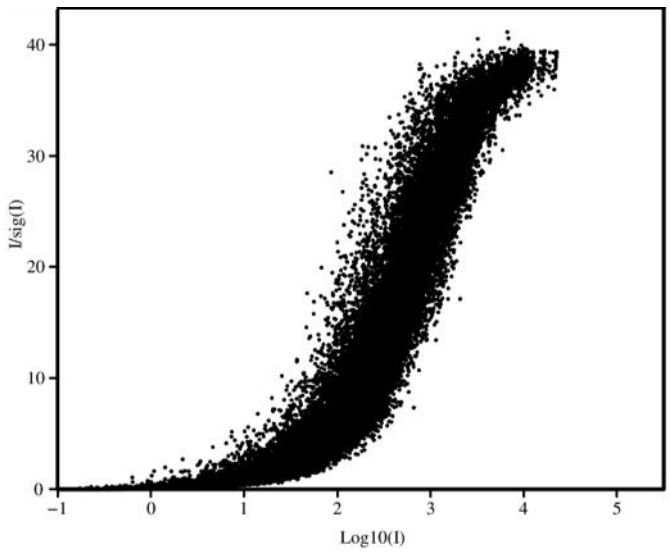


Figure 1.7 An example Diederichs plot[26] created with *SADABS*.

The I/σ_I or significance value for diffraction data is an indication of the accuracy of the measurements with the maximum possible value achieved being suggestive of the systematic error due to the experimental setup.[26] The maximum I/σ_I value also depends on

the error model applied to the standard uncertainties in the scaling stage. A graphical representation of the distribution of I/σ_I is provided by the Diederichs plot[26] (Figure 1.7). This plot, a graph of I/σ_I against log(I), is expected to have a sigmoidal shape. A plateau on the Diederichs plot suggests that a maximum I/σ_I limit has been reached and therefore increasing the exposure time of the measurements may not increase the I/σ_I ratio.

1.2.2 Descriptors of model fit

When a model of the electron density distribution of a sample has been created, comparisons between observed and calculated structure factors, F_o and F_c, can be made to quantify the degree of fit of the model to the data. The residual (R) factor can be used to demonstrate the agreement between observed structure factors and those of the crystallographic model. The subscript 1 indicates the R factor was calculated using F; a subscript 2 shows that squared structure factors (F^2) are used. R values are often given as a percentage – the closer the value of R is to 0, the better the match between the model and the experimental data.

$$R_1 = \frac{\sum\left||F_o| - |F_c|\right|}{\sum|F_o|} \qquad 1.13$$

The weighted R factor (wR) is also used, where w is the weighting function applied to a reflection.

$$wR_1 = \left(\frac{\sum|w\,|F_o - F_c|^2|}{\sum|wF_o^2|}\right)^{\frac{1}{2}} \qquad 1.14$$

Henn and Meindl[24] state that R_1 is a descriptor not a quality indicator as due to standard uncertainties of measurements an R_1 or R_2 of 0% is unlikely to be reached. Henn and Schönleber offer a measure of the predicted minimum R_2 that a dataset could reach, R_{pred} (Equation 1.15), to provide the expected degree of agreement with the level of noise (from σ) present in the data.[32] If R_2 is smaller than R_{pred} then it can indicate overfitting of the model – where the R factor of the model is reduced erroneously by fitting the model too closely to the crystallographic data. For further discussion of the identification and prevention of overfitting in the multipole model, see Section 1.4.6.

$$R_2^{pred} = \left(\alpha\frac{\langle\sigma^2\rangle}{\langle I_o^2\rangle}\right)^{\frac{1}{2}} \qquad 1.15$$

$$R_{meta} = \frac{R^{exp} - R^{pred}}{R^{exp}} \qquad 1.16$$

R_{meta}, the difference between R_{pred} and the experimental (or *de facto*) R_2 is given as a measure of the systematic error or 'mismatch' between the data and model (Equation 1.16).[24]

Another measure of the agreement of the calculated and observed structure factors, often quoted for crystallographic refinements, is the goodness of fit (GooF, Equation 1.17). This calculation is related to the χ^2 test in statistics, where it is used to calculate whether the differences between expected and observed frequencies are statistically significant.

$$GooF = \sqrt{\frac{\sum|\mathrm{F_o} - \mathrm{F_c}|^2}{\mathrm{N_{ref}} - \mathrm{N_{param}}}} \qquad 1.17$$

When the weight, w, is included in the value, the calculation is the weighted goodness of fit (wGooF), as shown in Equation 1.18. The sum of the squared difference of structure factors multiplied by the weighting scheme applied is divided by the degrees of the refinement (N_{ref}, the number of reflections minus N_{param}, the number of parameters being refined in the model). The ideal value for both the GooF and wGooF is 1.0.

$$wGooF = \sqrt{\frac{\sum \mathrm{w}|\mathrm{F_o} - \mathrm{F_c}|^2}{\mathrm{N_{ref}} - \mathrm{N_{param}}}} \qquad 1.18$$

The residual of each reflection is the difference between the calculated and observed structure factor. Henn reports that correlations between squared residuals and the standard uncertainty can result in a low wGooF value, regardless of the presence of systematic errors.[33] They state that the wGooF is only a valid measure if the standard uncertainties of reflections are adequate, the residual values are uncorrelated random values (which have a Gaussian distribution), and the squared residuals are independent from standard uncertainties. Henn suggests an alternative measure, aGoFs, which is not affected by this correlation (Equation 1.19). The aGoFs is expected to be close to 1 when there is no systematic error present in the data.

$$\mathrm{aGoFs} = \frac{\langle\left(\mathrm{I_{o,i}} - \mathrm{I_{c,i}}\right)^2\rangle}{\alpha\langle\sigma^2(\mathrm{I_o})\rangle} \text{ where } \alpha = \frac{(\mathrm{N_{ref}} - \mathrm{N_{par}})}{\mathrm{N_{ref}}} \qquad 1.19$$

A graph of F_o plotted against F_c is also useful in identifying possible outliers where the F_c shows a significant deviation from the global relationship to F_o. Other graphical methods include a plot of the variation of $\Sigma(F_o^2)/\ \Sigma(F_c^2)$ with resolution, which can identify any systematic trends in the data. The graph should be a straight line of zero gradient. The values are usually binned into resolution shells but can also be created for individual reflections to

identify any outliers. Both of these plots can be created using *DRKplot*,[34] a program created for use with multipole refinement software *XD*.[35]

1.2.3 Error distribution in multipole refinement

An important measure of the suitability of the refined multipole model is that the residual electron density is both flat and featureless.[19] The F_c from the multipole model is taken away from the F_o and a Fourier summation of the resultant values can create a representation of the areas within the unit cell where electron density has been under or over represented. A plot of the residual electron density containing large peaks or troughs of density indicates an inaccuracy or insufficiency in the model. 2D residual electron density plots are usually calculated for planes of the molecule that contain the most atoms, although 3D plots can also be created.[36] Figure 1.8 is an example of a 2D residual electron density plot; positive density occurs in regions where the multipole model has not modelled all the electron density present and negative density is observed where the multipole model provides an overestimation of the density, compared to that expected from the observations. The height of the maximum and minimum peak of residual electron density are often reported as a quality measure; if these are large they can indicate potential problems with the model or the data.

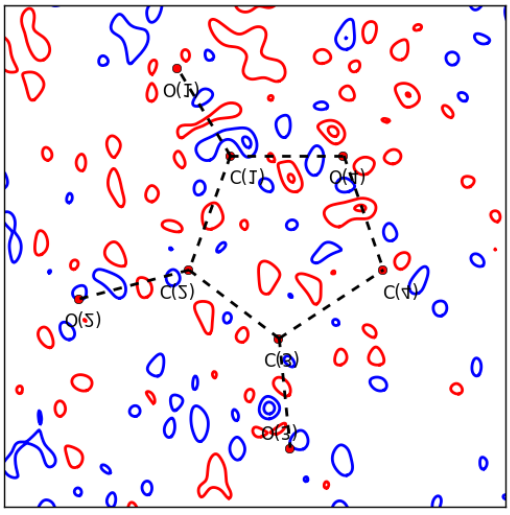


Figure 1.8 A residual density plot through 5 membered ring of vitamin C. Contours are at levels of ± 0.05e·Å^{-3} (negative density = blue contours, positive density = red contours).

As well as visually gauging whether the distribution of the residual electron density is flat or featureless, quantitative evaluation is possible using a series of descriptors formulated by Meindl and Henn.[37] The gross and net amounts of residual electron density can be calculated. The net residual density is determined by a summation of the value of residual electron density for a grid of a particular volume divided by the total number of grid points

analysed (Equation 1.20), where N is the number of grid points analysed and p_o is the residual density.

$$\rho_{net} = \frac{1}{N}\sum_{k=1}^{N} p_o(k) \qquad 1.20$$

The gross residual density is the total amount of residual density present in a particular volume of the unit cell (Equation 1.21). A scaling of ½ is included, as one misplaced electron results in a value of two electrons that are of place as there is increase of 1e in its incorrect position and a decrease in 1e from where it should be located.

$$\rho_{gross} = \frac{1}{2N}\sum_{k=1}^{N} |p_o(k)| \qquad 1.21$$

To calculate net and gross residual electrons, the investigated volume must be included (Equations 1.22 and 1.23).

$$e_{gross} = V\rho_{gross} \qquad 1.22$$

$$e_{net} = V\rho_{net} \qquad 1.23$$

These descriptors, e_{gross} and e_{net}, can be used to provide information about the flatness of the residual density. For a 'consistency check', the values of e_{net} and ρ_{net} must be close to 0 when the investigated volume is the unit cell. When the volume only covers part of the unit cell, the size of the value can indicate the amount of incorrectly assigned density in that volume. e_{gross} and ρ_{gross} do not vanish over the whole unit cell and describe inadequacies during the refinement process – the higher their magnitude the more electron density that is unaccounted for by the model. Henn and Meindl demonstrated the application of e_{net} and ρ_{net} to real and synthetic data and how they can be used in identifying systematic errors in a refinement model or dataset in their 2008 paper.[37]

Alongside descriptors of the flatness of the residual electron density, Henn and Meindl describe a measure of the featureless-ness of the distribution of residual electron density. Their proposed method uses fractal dimensionality analysis. The residual electron density distribution can be considered as a fractal and by calculating the dimension of the iso-surface of the residual electron density at 0.01 eÅ^{-1} intervals a fractal dimensionality plot can be created and compared between different structures. The ideal distribution is parabolic (Figure 1.9) and deviations from this shape imply errors other than that of experimental noise. The effect of systematic errors on the fractal dimensionality plot is reported by Henn

and Meindl, with some systematic errors identified by their characteristic features within the fractal dimensionality plot.[37] The reporting of fractal dimensionality plots has recently become more commonplace within literature, but there is no requirement and their usage is not universal throughout the field. Wolf *et al.* stress the importance of using fractal dimensionality plots to assess refinements to identify potential problems with refinements.[27]

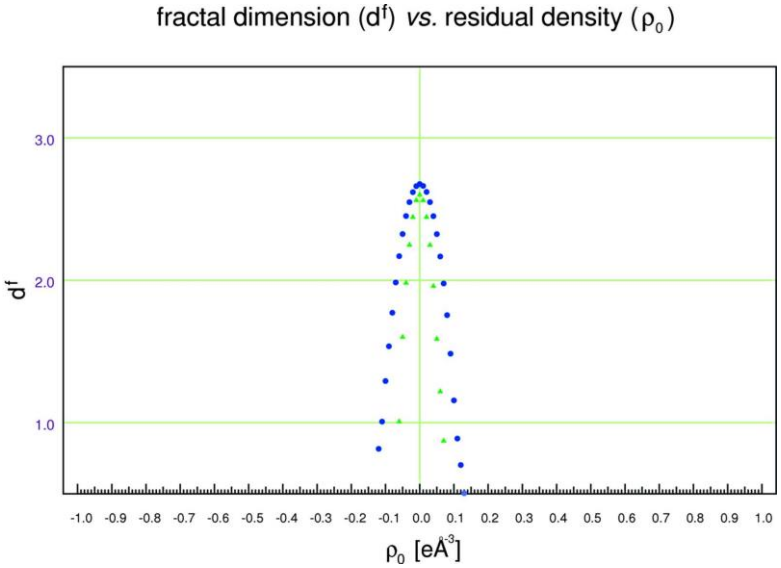


Figure 1.9 A fractal dimension distribution for simulated data with a low level of Gaussian noise, recreated from Henn and Meindl[37]with permission from Acta. Cryst. A. The plot shows effect of truncation of data (green triangles $\sin\theta/\lambda$ cut-off 0.80 Å and blue circles 1.14 Å).

If no systematic mismatch occurs between the data and the model, the distribution of residuals from a least squares refinement (not to be confused with residual electron density) in a multipole refinement, calculated from the difference of F_o and F_c (Equation 1.24), should be random.[32] Therefore, the residuals of any model should be normally distributed – provided the structural model is correct and the applied weights and any errors in the measurements have been evaluated correctly.

$$\text{Residual} = \sqrt[2]{w}(F_c^2 - F_o^2) \qquad 1.24$$

Normal probability plots appear frequently alongside charge density refinements as a graphical measure of the distribution of residuals and way to determine if the structure factor residuals in a charge density refinement have a normal distribution.[24,38,39] If this is not the case, then there may be systematic errors present in the model or the data. The distribution of the structure factor residuals is described as being normally distributed if the normal probability plot has a gradient of 1 and an intercept of 0. Like residual density analysis, there have been repeated calls for normal probability plots to be included alongside charge density models as a descriptor of data quality but again this practise is yet to be universal.[27,38] Henn and Meindl report that as not all published charge density refinements

have a normal distribution of residuals, those without should therefore be treated with caution.[40,41]

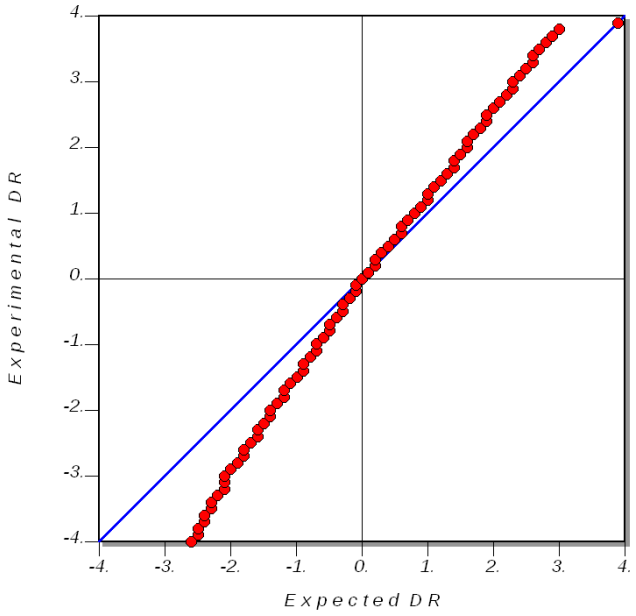


Figure 1.10 A normal probability plot created with DRKplot.

DRKplot is a commonly used program for creating normal probability plots for charge density data. [34] In this program residuals are binned in a range of + 4 and – 4 according to their experimental value and an average is plotted. In reality, this range can be a fraction of the total residual values. Figure 1.10 shows a normal probability plot created with *DRKplot* for a model which does not have a normal distribution of residuals.

To identify systematic correlations between residuals and different reflection properties in the data, Henn and Meindl have proposed a variety of other techniques including Bayesian Conditional probability (BayCoN) plots,[42] graphs of the mean binned residual density, and 'rare-event' plots across properties such as intensity and resolution of data[41] to indicate when correlations have occurred. Residuals correlated to a property of the reflection are by definition not truly randomly distributed and therefore may signify the existence of systematic errors within the data.

1.2.4 Descriptors of model reasonableness

The *XD* manual suggests that the model should also be evaluated by its ability to recreate expected topological features e.g. BCPs in predicted places. If this is not the case, it could indicate that the model is substandard.[43] Other properties, such as the Laplacian of the electron density, are said to be very sensitive to the concentration of charge within the model and should thus also be scrutinised.[43]

Deformation electron density maps can be created to view the distribution of electron density within 2 or 3D in the unit cell. There are two main types: dynamic and static. Dynamic maps are calculated including the ADPs, which can cause thermal smearing of the parameters resulting in less fine detail being discernible in the maps. These are calculated through the Fourier summation of the difference of calculated multipolar structure factors and calculated spherical atom structure factors (similar to creation of residual electron density plots). Static deformation density maps, in contrast, are calculated in direct space and do not contain any information from ADPs.[19]

The Hirshfeld rigid-bond test can be used to check if the ADPs for the model are physically reasonable and not biased by vibrational smearing.[44] For covalently bonded atoms of carbon and heavier, it is expected the difference in the mean-square displacement amplitudes of bonded atoms along the direction of the bond is ≤ 0.001 Å^2. This test is based on the theory that bond stretching contributions to vibrations are smaller than other contributions to the overall vibrational motion of the atoms, therefore, by assuming the bonds are rigid, the relative vibration motion should vanish along the bond. This in turn leads to the expectation that the radius of the ADPs direction of a bond should be approximately equal. Larger than expected mean displacement amplitudes can suggest ADPs biased by systematic error or the existence of more complex modes of vibration within the molecule.[45]

The Shapiro-Wilk test[46] has also been used to assess if parameters across refinements of the same sample have a normal distribution.[47] This is a hypothesis test to assess whether populations are normally distributed. The null hypothesis, that the data are normally distributed, is rejected if the p value for the data is below a certain significance or α level. An α level of 0.10 means there is a 10% chance of rejecting the null hypothesis incorrectly. This test is applicable to datasets containing 3 to 5000 values.

1.3 Data files created during experiment

There are a number of different files created during the experimental process, from data collection through to structural refinement, containing information about the experiments which will be referred to throughout this thesis. The formats of the files output during the process of an experiment and refinement are specific to the manufacturer and programs used, but the overall information provided by each file is similar across the field.

- Diffraction frame files – the files containing the experimental diffraction pattern as measured by the diffractometer. They comprise of the intensity measurements from the detector, alongside a header which holds the experimental metadata e.g. temperature, instrument type, position of the detector and setting angles when recording the information. Diffraction frame and image are often used interchangeably – in this thesis frame refers to the entire file and image refers just to the intensity information recorded from the detector contained within the frame.
- *.raw* – the output from the integration program *APEX*/*SAINT* containing information for each reflection including where it was measured on the detector, the run number, and any corrections applied in the integration stage. It also shows the initial corrections which are applied to the file. Further details on the contents of the file can be found in Section 10.7.
- *.hkl* – a processed list of all the recorded reflections and their standard uncertainties, usually created during the scaling stage of refinement. This file can be merged, where all reflections which are expected to have the same value are averaged to give one intensity and standard uncertainty value, or unmerged, where information for each individual reflection measured is recorded separately.
- Structure factor files (*.fcf* or *.fco*) – files created during the structure solution and refinement and contain the calculated and observed structure factors, alongside the standard uncertainty of each reflection.
- Crystallographic Information Format (CIF) file – these are required files for most journals when publishing crystallographic data. They contain information about the experiment (temperature, corrections applied), the crystal structure (including unit cell parameters), and increasingly the measured structure factors.

1.4 Considerations for a charge density refinement

The details of the method for processing data should not be seen as a one size fits all approach. For charge density there is no absolute method agreed upon within the field, rather a choice of different methods and programs that can be employed which all operate in slightly different ways. Even within the individual programs, there is a choice of different options that can be utilised such as the corrections applied, cut-offs for the data to be excluded from the refinements, and different algorithms for the processing of data – the

combinations of those applied vary across published data. This variance implies an uncertainty in approach and a potential large source of error within the literature.

1.4.1 Sample and crystal selection

While the decision of what sample to select is not a variation of the data processing, the choice is nonetheless very important. As the multipoles are based on the valence electron densities, a greater number of core electrons leads to a smaller ratio of valence:core contributions to scattering at low angles (Figure 1.11) which can make charge density refinements more challenging.[48] Stevens and Coppens suggest a suitability factor for samples for charge density studies;[48] the suitability factor, S, is calculated from Equation 1.25, where n is the number of core electrons and V is the volume of the unit cell.

$$S = \frac{V}{\sum n_{core}^2} \qquad 1.25$$

This factor would suggest that the best samples for charge density refinement are those with few core electrons i.e. those containing first row elements with small unit cell volumes, however, charge density studies are increasingly performed with transition metals[49] and larger molecules.[50–52]

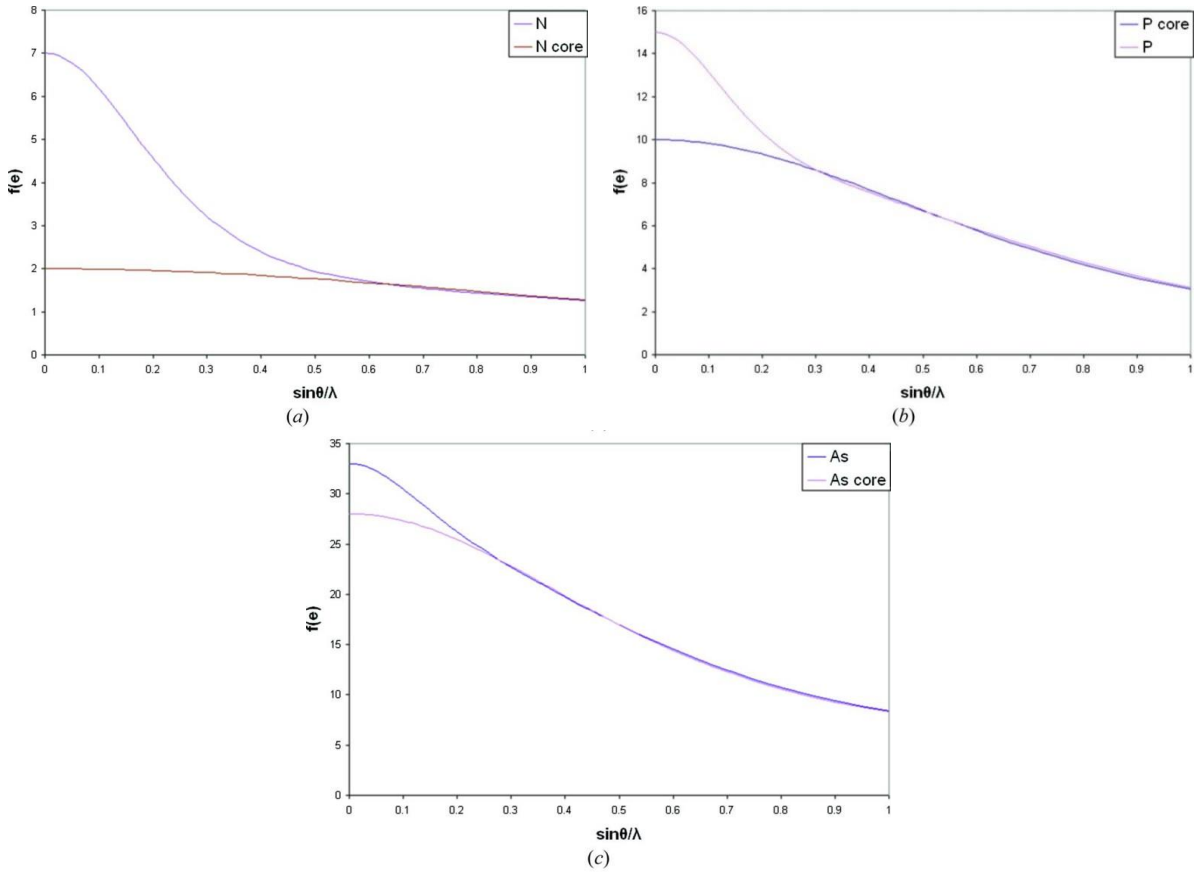


Figure 1.11 Total and core scattering for elements N, P and As as a function of sin θ/λ (Å^{-1}), reproduced from Zhurov et al.[38] with permission from J. Appl. Cryst.

After the choice of the sample, another important factor in obtaining data for a successful multipolar refinement is the quality of the crystal selected from a given sample. Destro *et al.* illustrate this point for charge density refinements in a comparison of two different high resolution, low temperature, datasets they labelled medium and high quality (1.14 $Å^{-1}$ $(\sin\theta/\lambda)_{max}$, 19.5/19.0(10) K).[53] The medium quality dataset has both a reduced reflection redundancy (4.74 vs 1.1 ratio of mean measured/independent reflections) and a poorer quality crystal, which produced reflections *"tainted by little imperfections"*, including *"extraneous peaks"* in reflection profiles, visible from close analysis of the recorded data (Figure 1.12). When choosing the crystal to use in the data collection stage, it is generally accepted that a good crystal for charge density refinements is one that is single and is optically pure i.e. has no defects or cracks.

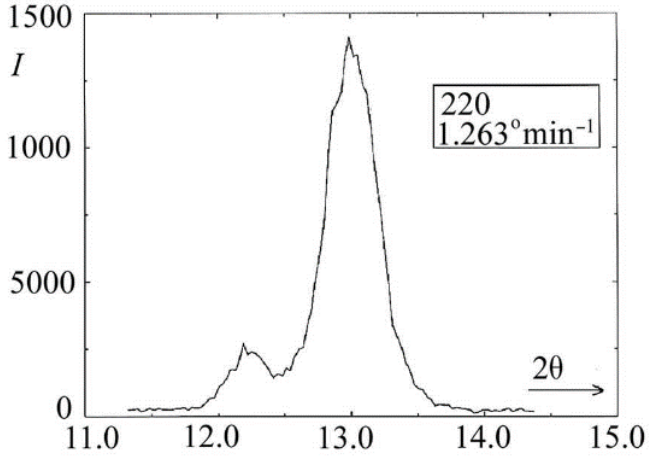


Figure 1.12 Reflection profile with 'extraneous peak' as described by Destro et al.,[53] reproduced with permission from Acta. Cryst. A.

An analysis of resulting multipole refinements and the properties gained from them by Destro *et al.* suggested that while the overall molecular properties derived from the models are in agreement between the medium and high-quality data, quantitative analysis of electron distribution should be reserved for only *"genuinely high quality"* data. Plots of the derived Laplacian are used to highlight differences in electron distribution; most notably, features present in the Laplacian derived from the higher quality data, such as polarization of hydrogen atom density in hydrogen bonds, are absent in the Laplacian of the medium quality data. Lower standard uncertainties for refinement parameters for high quality data (2-3 times smaller) were reported, which was attributed to greater precision of measurements. The higher quality dataset was also found to have lower refinement statistics (R_1, wR_2).

1.4.2 Data collection

Careful data collection is a necessity throughout all scientific investigations and a prerequisite for obtaining the highest quality data. Within the field of charge density, it is of paramount importance as multipolar parameters are calculated from the scattering associated with the valence electron density, which is often only a small portion of the total structure factor. A number of different choices are available to the researcher during data collection – although some may be dictated by the experimental setup (such as detector type and X-ray wavelength available). Other default parameters require significant attention to obtain the data required for a charge density study.

Thermal diffuse scattering (TDS) is a source of systematic error,[5] which occurs due to atomic vibrations within the crystal.[54] This effect reduces with temperature – charge density data collections are almost exclusively performed at 120 K or below, using nitrogen or helium cooling devices. TDS can cause increases in the intensity measured for reflections and affect the ADPs.[55] A multi-temperature study using single crystals of naphthalene showed that the thermal motion could be deconvoluted from other parameters in a multipolar refinement, as there was little change in the multipole parameters when the temperature decreased.[56] Resolution dependant scaling was used by Niepötter *et al.* to investigate whether datasets required a correction for TDS.[57] There have been a number of corrections developed for TDS that are empirical,[54] but their use is far from commonplace.

The method of the generation of X-rays for use in the diffraction experiment is also a consideration. There are a number of different convenient wavelengths that can be utilised for X-ray diffraction in the home laboratory: the most common available are copper (Cu, λ = 1.54184 Å), molybdenum (Mo, λ = 0.71073 Å) and silver (Ag, λ = 0.56086 Å). The different metals have unique X-ray emission profiles leading to peak intensities at different wavelengths.

The main differences between the respective wavelengths are the absorption and scattering intensity of the crystal being tested, along with the changes in the Bragg angle for the given reflections. Decreasing the wavelength from Cu to Ag results in less absorption of X-rays by the crystal but also less intense reflections being produced (when measured for the same amount of time) due to the difference in energy of the X-ray photons. A shorter wavelength also leads to a reduction in the θ angle a particular reflection is measured at, which can cause an increase in the available resolution that can be measured by a particular

experimental arrangement. The highest possible resolution, d_{hkl}, a detector can measure data at is limited by the maximum 2θ angle a detector can reach without collisions. The distance between reflections as measured on the detector surface (provided detector distance remains the same) will also decrease – which can be extremely advantageous in other techniques where the available data collection volume is limited.[58] A comparison of Mo and Ag sources using the spherical atom model suggested that the use of Ag radiation resulted in models with better refinement statistics, especially for strongly absorbing crystals.[59]

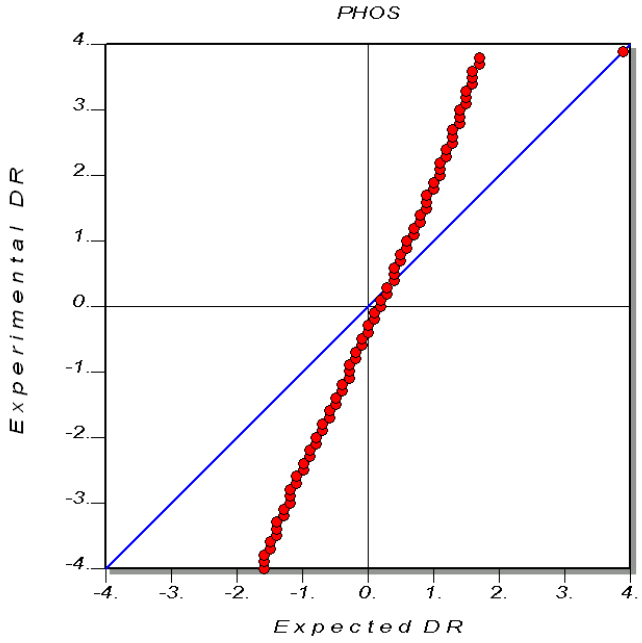


Figure 1.13 Normal probability plot with underestimated standard uncertainties. This figure is reproduced with kind permission from Dr. J. Henn.[40]

Multiple measurements of individual reflections are required to improve the standard uncertainty of the measurement of the reflection; the number of measurements of each reflection is defined as its multiplicity. This is different from the redundancy of the reflection, which requires the measurements of each reflections to be unique (i.e. measured at different positions on the detector). Highly redundant data are important to ensure accurate determination of reflection intensities[53] and their associated standard uncertainties.[26,60] However, as a consequence of high multiplicity the standard uncertainties can be underestimated,[61] which was shown with the use of synthetic data to have a characteristic shape in the normal probability plot (Figure 1.13).[40] Standard uncertainties are calculated in a variety of different ways as described in Section 1.4.4.

The ratio of unique reflections measured to the parameters used in the refinement is important; this figure guides the user when deciding on the number of model parameters to

introduce. The standard accepted value for a spherical atom refinement is > 10. How relevant this value is to charge density data has been debated[19,62] (and is elaborated on in Section 1.4.6). However, to ensure this arbitrary ratio is reached for multipole refinement and the additional parameters that are required, diffraction data are collected to a higher resolution than would often take place for a spherical atom refinement (≤ 0.5 Å).[19]

Higher resolution data can also allow a more accurate determination of atomic positions. It is possible to see that in Figure 1.11, along with drop of the structure factor with sinθ/λ (Å^{-1}), the scattering power of the valence electron density decreases much faster than that related to core electrons. Therefore, at high angles of 2θ or high resolution, the structure factor information comes almost solely from core electrons, which is further illustrated in Figure 1.14. Refinement of atomic positions against high 2θ angle reflection data is sometimes used to gain more accurate locations of the atoms.

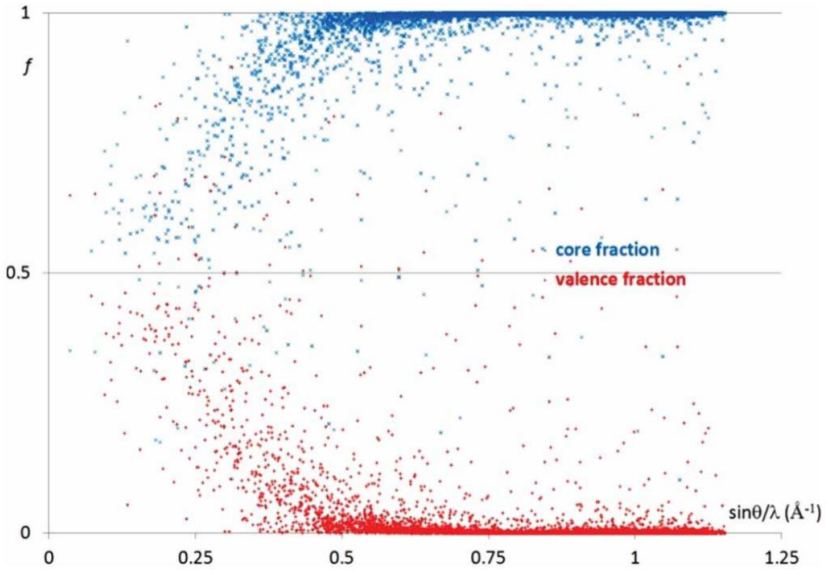


Figure 1.14 Fraction of core and valence densities contribution to total reflection intensity for L-histidinium hydrogen oxalate, reproduced from Macchi[19] with permission from Taylor and Francis.

Other considerations include the width of the scan included on each frame and the length of time taken to measure each scan, which can be varied during a data collection. Sørensen and Larsen investigated the effect of scan width and exposure time using a charge coupled device (CCD) detector.[63] A single crystal of tetrafluoroterephthalonitrile was measured using scan widths of 0.5° and 0.25° collected with an exposure time of both 70 and 140 s per frame. An additional dataset with 1.0° frames with a 70 s exposure was also collected. It was reported that for a decreasing scan width, the R_{int} decreased by a small amount (2.54 to 2.40 % from 1° to 0.25°). However, the standard uncertainty of measurements increased with the 70 s scans. Increasing the length of the scan exposure time decreased both the standard

uncertainty and the R_{int} values (0.5° – 2.45 to 2.14 %, 0.25° – 2.40 to 2.20 %). Charge density refinements were not completed on any of these collected datasets, therefore unfortunately the effect of the collection on the final multipole models can only be inferred from this study.

Increasing the exposure time of the diffraction frames increases the total collection time. As charge density datasets are often collected to a high redundancy, requiring a large number of runs collected at different setting angles, data collection times of the order of days are still commonplace to achieve the highest quality data. Therefore, the decrease in standard uncertainty that could be achieved through longer exposures must be balanced with the length of instrument time available, along with crystal and instrument stability.

In any research laboratory there will be certain experimental restrictions to a given user. This is particularly true in relation to the detection device employed, unless there is an availability of multiple different detectors. Currently small molecule diffraction primarily uses CCD detectors and complementary metal-oxide semiconductor (CMOS), although in the past 10 years further development has been made to produce hybrid pixel detectors (HPAD) and charge integrating detectors (CPAD) compatible with small molecule crystallography. Each operates in a slightly different way, but all are primarily made up of a 2D grid of cells (or pixels) which can detect electromagnetic radiation.

CCD

In general, CCD based detectors are made up of three main parts: a phosphor, a coupling method and an imaging array. The method of X-ray detection by these detectors is described briefly below; for a more detailed account see Gruner *et al.*[64]

X-ray photons are not able to be measured directly by a CCD chip, therefore, the number of X-rays that hit the detector at a particular point are measured indirectly through the conversion of the X-ray photons to visible light. Conversion is achieved through use of a phosphor or a scintillator, both terms are often used interchangeably within the literature to describe the same item: a material which emits visible light upon being struck by X-rays. The phosphor material is crafted into a thin uniform layer across the front of the detector.

The visible light then needs to reach the imaging array. Fibre optic bundles are used in many detectors in order to direct photons towards the imaging array and increase the size of the detection plane. The low refractive index of the cladding of the fibre as compared to the

core causes total internal reflection to occur at most angles that a photon hits the cladding. This prevents the photon from continuing to travel in their given direction, guiding it towards the imaging array.

CCD chips are made out of a semiconductor which converts photons of visible light into electrons via the photoelectric effect. A potential well is formed by applying a voltage to an electrode below the semiconductor. This voltage attracts the electrons produced and holds them in place until they are measured. The number of electrons stored is proportional to the number of X-ray photons that hit the detector at any given point.

Reading the number of electrons in each well and converting them to a digital format takes place one pixel at a time. There is a reader at the corner of the imaging array – electrons are shunted sequentially into the end column and then shunted down the column to the reader. The reader converts the charge in the readout well into a voltage, which is then measured, and the value stored. This process takes time during which the detector can't collect information, often called readout or 'dead' time. More advanced detectors have a readout chip at each corner of the pixel array to speed up this process and reduce dead times.

Since their development, there have been studies probing the suitability of CCDs for collecting the high resolution diffraction data, that is required for charge density studies.[65] Compared to image plate detectors, their predecessor, CCD data was found to be of a comparable quality with the added benefit of faster experimental times. However, CCDs are not without their problems and have been found to underestimate the standard uncertainty of reflections – a factor crucial in getting the best structural models for charge density data.[40,66]

CMOS

CMOS detectors work in much the same way as CCDs detectors, however, unlike CCDs, CMOS arrays have a readout chip on each pixel that enables the measurement of the number of electrons in its potential well. The readout from these chips is therefore much faster;[67] this allows a 'shutter-less' continuous rotation method of intensity measurement. Instead of requiring the X-ray shutter to be closed while the readout of the intensity of each pixel takes place, as with a CCD, the readout time for each chip is sufficiently short that the sample is continuously rotated as the detector records the diffraction pattern without pause.[68] Data from this type of detector have not been used in this thesis.

PAD

Unlike CCD and CMOS detectors, HPAD work by counting each photon that hits the detector individually.[67] They consist of a sensor layer, which converts the X-ray photons to electrons without the requirement of a phosphor, and an array of readout chips. The lack of a phosphor generally leads to a lower point spread function compared to CMOS and CCD based detectors, where the internal scatter of X-rays during their transmission through the phosphor results in a wider spread of the intensity relating to that reflection on the detector. When a detector measures a charge greater than a threshold amount, one single photon is counted. This leads to a low detector noise,[69] as charge measured below the counting threshold of the detector is simply ignored. Intensities for each pixel are measured in a CMOS style, with a readout chip on every pixel.

However, HPADS also have associated problems – the most well-understood being count rate saturation. The measurement of an X-ray photon takes a certain amount of time, therefore photons striking the pixel while it is recording will not be measured. There is also the issue of charge sharing between pixels, where an X-ray photon strikes the detector close to the boundary of an individual pixel. When this occurs, the charge created can be 'shared' between neighbouring pixels. This not only causes a reduction of the counts measured for an individual pixel but can cause some photon counts to be lost due to their incorrect identification as noise.

The pixel size for the HPAD is also much larger than CCD/CMOS – 0.172 mm for Dectris Pilatus 300K vs 0.120 mm for APEX2 CCD detector. This can result in intensity for a reflection being measured on just a few pixels on the HPAD detector. A small point spread function of a reflection can create problems within the integration process, causing difficulties in certain integration algorithms that fit a profile to a reflection in order to measure the intensity – especially in cases where strong reflections cover just a few pixels.

A very recent development in the field is the CPAD, which has been claimed to circumvent the problem of charge sharing and count rate saturation that affect HPADs.[70] A CPAD measures the full charge created when X-ray photons strike a pixel on the detector within a given time, instead of single photons as in the HPAD. Accurate voltage measurements allow the identification of the number of photons which have struck each pixel.

Nevertheless, there are problems that exist for CPADs. As with CCDs, saturation of the detector can occur if more charge than the potential wells can store is measured, resulting in a loss of intensity measurement.[70] Also unlike HPADs, the Photon II CPAD created by Bruker AXS Ltd., upon which most of the data in this thesis was collected, does contain a phosphor to convert X-rays to visible light.[70]

1.4.3 Integration

Integration is the process where an intensity and associated standard uncertainty is extracted from the raw experimental data for each reflection. There are a number of different software packages for the integration of crystallographic data, both commercial and not for profit. Integration programs often offer many different corrections that can be applied when processing diffraction data. The corrections that are available, and the form of those corrections, depend on the program being used. Additionally, there are different algorithms that can be employed to measure the magnitude of a reflection.

Principally, there are two main approaches that have continued into common usage: simple summation and profile fitting. These algorithms are utilised in a variety of ways within different packages. An outline of these is given below, however, for more detailed description of the algorithms see the documentation of a particular software.

Profile fitting assumes that all reflections have a similar shape in reciprocal space and intensity distribution when they are converted to precession-like geometry.[9] A 'standard' profile is determined, which is described by Leslie.[71] In the *SAINT* integration program,[72] which is used in this thesis, this profile is based on the shapes of the stronger reflections - the default setting is those with $I/\sigma_I > 10$; depending upon the algorithm, this profile can be in 2 or 3D.[73] The total intensity of a reflection can then be calculated by fitting the intensities observed to the standard profile. This method has been proven to be useful when evaluating the intensity of weak reflections,[71] where data are more strongly affected by the presence of detector noise. This algorithm has been shown to improve the refinement statistics for data with many weak reflections.[74]

In the simple summation algorithm, the total intensity of a reflection is not determined by fitting measured intensities to a profile, in contrast to profile fitting. Firstly, the pixels that contain intensities that contribute to the reflection are identified. These intensities, after the background has been subtracted, are added together to generate the intensity of the

reflection on a particular frame.[12] If thin slicing techniques have been employed, where intensity from a reflection appears on multiple consecutive frames, this value is summed with the values of the reflection appearing on consecutive frames to form the total intensity value.

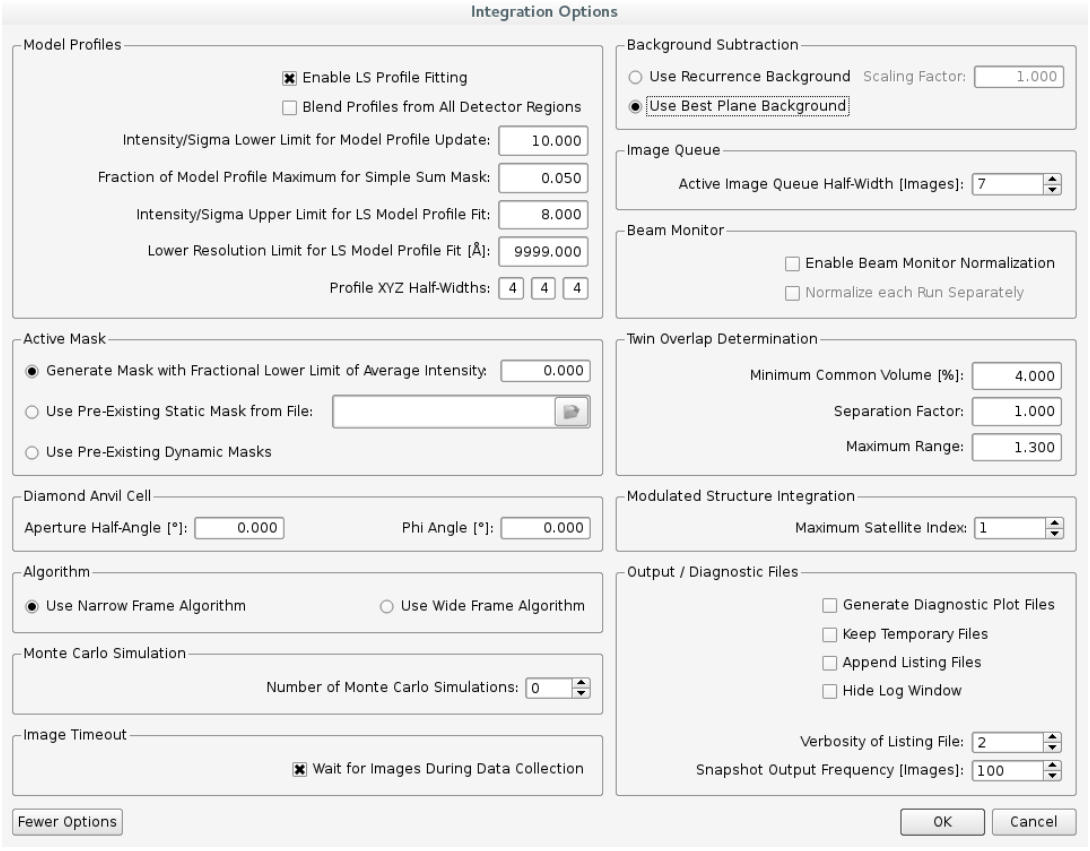


Figure 1.15 Screenshot of integration options screen as available inside the APEX3 GUI (which uses SAINT).

An image of the integration options screen within the *APEX3* graphical user interface (GUI),[75] which uses the *SAINT* program to perform the integration, is shown in Figure 1.15 to demonstrate a number of the choices available to the user. In *SAINT*, the algorithm with which to measure the intensity of the reflections can be chosen. The default setting is to use a combination of both algorithms; if the I/σ_I of a reflection is less than 8, it will be measured using the profile fitting algorithm, while reflections above this threshold are measured with the simple summation algorithm.

Contributions from experimental noise give background intensity across the detector, which must be removed in the calculation of reflection intensities. This background can be calculated in a variety of ways; within the *APEX3* software, there is the choice for a recurrence or a best-plane background. Additional corrections can be edited through operation of the integration package from the command line, giving the option to turn off or change default corrections for the detector (e.g. phosphor efficiency).

1.4.4 Scaling

To combine data from different experimental runs, corrections must be applied for additional systematic errors that were not taken into account in the initial integration measurements to make reflection values from different runs comparable. This is known as scaling. The aim of these procedures is to massage the data through the addition of corrections, so that the intensity value of identical reflections that have been measured in different experimental runs are more similar to each other. The data can also be merged, which is required for charge density, meaning that the individual measurements of each reflection are combined to give one value for each reflection. The method of merging, scaling and applying corrections varies between programs and approaches.

SADABS is an example of a popular scaling program.[28] The program consists of two stages, first the parameter refinement, where the intensity values are scaled, and then an error model refinement, which is applied to the standard uncertainties.[76] The parameter refinement involves several corrections which are applied to the intensities (Equation 1.26). A scale factor, S(n), for each frame is calculated which is constrained against the value of the scale factor for the previous frame. A correction for the absorption of the beam by the crystal, P(u,v,w), is also applied. This can be either numerical or analytical.[77] A correction for absorption based an accurate model of the shape of the crystal – by locating the crystal faces from a video of the crystal – can also be applied.[78]

$$I_c = I_o S(n) P(u,v,w) \qquad 1.26$$

A weighting scheme (Equation 1.27) is applied to the parameter refinement. A value for the weighting scheme parameter, g, can be defined by the user (default value = 0.04); this value is multiplied by the mean intensity of the corrected reflection and is refined later in the error model refinement. Robust/resistant least-squares refinements are used to optimise the values for all the corrections applied in the parameter refinement.[76]

$$w = (\sigma^2 + (g\langle I_c \rangle)^2)^{-1} \qquad 1.27$$

After parameter refinement, the error model of the standard uncertainties is then refined. Reflection measurements are deemed to be outlier and rejected if the intensity value is not within the range of $\pm n\sigma_I$ from the average value of the intensities (default value is $\pm 4\sigma_I$). The error model, shown in Equation 1.28, is then applied to the uncertainties. The software optimises the combinations of K, a scaling factor, and g, an intensity dependant correction sometimes known as the experimental error,[76] to ensure the weighted mean square

deviation χ^2 is as close to 1 as possible across *"the full ranges of intensities"*.[59] The default error model within SADABS is to apply an individual value for K that is optimised for each run, with a value for g that is optimised for the whole experiment. It has been suggested that the default error model is not the most suitable for charge density refinement;[79–81] there are other error models where it is possible to use only one of the K or g values, refine an individual value for g for each run, or manually set these values within the program.

$$\sigma^2 = K(\sigma_c^2 + (g\langle I_c\rangle)^2) \qquad 1.28$$

SADABS also produces a variety of output plots (such as scale factor for each frame, R_{int} with resolution and a Diederichs plot) to give information about the scaled data. The corrected intensities and standard uncertainties alongside the reflection indices are output in *.hkl* files. *SADABS* can also output merged *.hkl* files in the format required for *XD*, with the option to remove systematic absences and negative intensities (as these are ignored in *XD* refinement).

SORTAV[82] is another commonly used scaling program that works in a very similar way to *SADABS*. However, in this case the standard uncertainties are calculated taking into account the weighted root mean squared measurement of the errors propagated through from estimates of experimental errors (external error) and weighted root mean square value of measurements from their mean value (internal error).[60] *SORTAV* also scales the intensities of reflections in batches as opposed to each frame individually, as in *SADABS*.

Henn and Meindl reported on the effect scaling programs had on the maximum significance, I/σ_I, of observations.[79] The treatment of data and the application of corrections can either leave the standard uncertainty unchanged or increase the value, thus decreasing the significance. This is due to the addition of further uncertainty of the measurement when additional parameters and corrections are applied, causing errors to be propagated throughout the course of data processing. A smaller drop in significance of reflections appeared to be observed when processing data with SORTAV as opposed to *SADABS*, although no comment was made by the authors as to the relative reduction in significance.

Jørgensen *et al.*[80] continued these investigations and compared the significance of reflections when scaled using *SORTAV*, *SADABS* with the default error model, and *SADABS* without an error model applied to the standard uncertainties against those from the raw data before background subtraction (Figure 1.16). This showed the significance of the data

scaled using the *SADABS* default error model reached a plateau (red plus signs on Figure 1.16), which could be a function of applying an intensity dependant correction to the standard uncertainty. No comment by the authors was made at the apparent increase in significant at low values of I/σ_I, a feature that was especially apparent for *SADABS* scaled data. Data scaled with *SORTAV* (blue stars) remained almost as significant as the raw data. This was the only example given, therefore it is not possible to ascertain if this trend is present across all data sets.

The authors report that effect of error estimation in *SADABS* does not have *"significant effect on final model of charge density data"* but evidence to support these claims was not provided. Henn and Meindl,[81] in response to Jørgensen *et al.*, state that although the calculation of standard uncertainties using *SADABS* does not have a significant effect on the final model parameters, there still may be an effect on the topologies of both the Laplacian and the density calculated from the model. Henn and Meindl recommend carrying out data processing in such a way that as little significance as possible is lost.

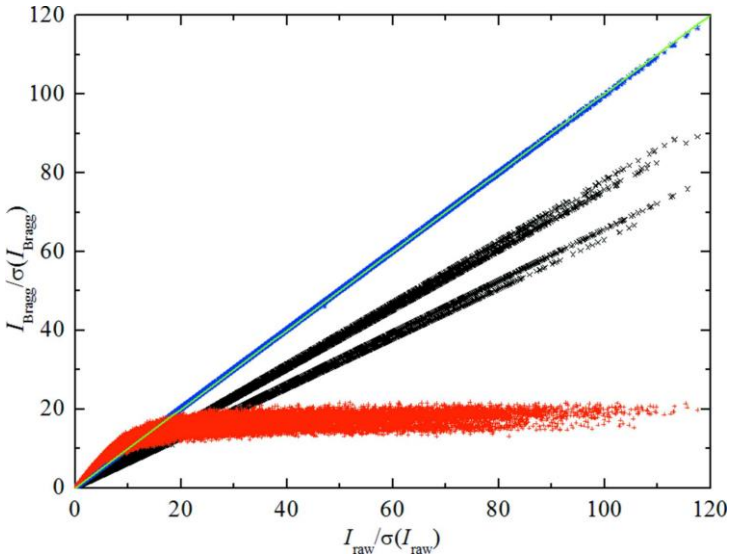


Figure 1.16 A comparison of significance of raw intensities against Bragg intensities processed using different software methods, reproduced with permission from Jørgensen et al.[80] *(Key of values: blue - SORTAV, red – SADABS + error model, black – SADABS no error model).*

Another process that has been utilised with charge density data is 'resolution dependant' scaling. This was discussed in Section 1.4.2 as a way to validate if a correction for TDS is required.[57] Data were divided into ten separate bins of 0.1 $Å^{-1}$ increments and then each scaled separately using the *SADABS* default error model so an individual scale factor, K, was calculated for the data in each bin. If the scale factors for each separate bin were not approximately the same, this was said to suggest the need for a TDS correction (as TDS is a resolution dependant effect). Niepotter *et al.* show that applying resolution dependant

scaling can reduce the differences in residual density, making a 100K data collection more similar to a 15K data collection of the same crystal.[57]

1.4.5 Structure solution and refinement of the spherical atom model

During this stage of the data processing, an approximate model of the electron density is created, which is then iteratively improved upon during the refinement of the structure. Within both of these stages, there are very few choices that can be made. This stage is preparation for applying the multipole model; the input for *XD* requires the spherical atom model solved with atomic positions and the *.hkl* file. The completed spherical atom model is then converted into input files to perform the multipolar refinement.

1.4.6 Multipole model

There are various different software packages that can perform multipole refinements. In this thesis *XD2016*[35] was chosen, due to it being developed for small molecules. This suite consists of a variety of programs for performing different stages of the multipole refinement and subsequent analysis of the data.

When importing the data to *XD2016*, using the program *XDINI*, there is a choice of four databanks which can be used to define the atomic wavefunctions. The form of the different databanks are described in the *XD2016* manual,[43] which is distributed alongside the program, differ between each databank and may result in differences in the form of the multipole density. Files from a spherical atom refinement are taken and converted into the format required for refinement with *XD2016*.

The least squares refinement in *XD2016* can be performed using either F or F^2 – both have advantages and disadvantages.[83] Watkin elaborates that F can become a problem when assigning standard uncertainties for small or negative values of intensity,[84] causing problems when weighting schemes are applied. Also, if the F_o and F_c do not have the same sign then there is an issue with having a continuous residual density. Additionally, the refinement can be affected by large errors in weak data. Conversely, F^2 refinements are affected by large errors in strong data. It was suggested, in a teaching review of data processing with the spherical atom model by Müller,[85] that F^2 is the superior measurement to use in the least squares refinement as it is possible to include negative intensities in the refinement (which is more complicated when refining against F). However the default cut-off for intensities applied in the *XD2016* least squares refinement package (*XDLSM*) is that reflections with

negative and zero standard uncertainties are removed, therefore this may not be a consideration.[43] When merging data with *SADABS* for use in *XD* these reflections, along with those of I < 0, are suggested to be removed from the *hkl* file.[76]

When performing a multipole refinement, different levels of multipole parameters (e.g. monopoles, dipoles etc.) are usually added to the refinement sequentially. Which multipolar parameters to include for different atoms within the molecule is an important consideration when undergoing a refinement. Generally, hydrogen atoms are only refined up to using the monopole level and one dipole included along the direction of the bond to the hydrogen; non-hydrogen atoms are refined up to and including octupoles for elements in the first row of the periodic table and up to hexadecapole level beyond this. Adding in higher levels of multipoles means more parameters and can also cause overfitting of the data.[62]

Cross-validation has increasingly been used in multipole modelling to try and prevent the overfitting of data.[86] In a cross-validation refinement, reflections are divided into two different subsections of different sizes; most reflections belong to the larger section, called the work section, which will be used in the refinement of the multipole model. The other reflections belong to the free section. Reflections in the free subsection are not used in the refining of the model and only to calculate data quality indicators. The R_1 value calculated with the refined model using the free section of reflections, also known as R_{free}, shows how well the model fits to reflection data that have not been used in its creation.[87] If adding parameters into the refinement of the work section results in a drop in R_1 but there is not a corresponding drop or there is an increase in R_{free} then the model may be overfitting the data.

Constraining parameters in the refinement is one way to lower the risk of overparameterization. This has the effect of reducing the number of parameters to be refined and thereby increasing the data to parameter ratio. Constraints on the multipole populations of atoms in chemically similar environments can be applied within a multipole refinement – keeping the multipole populations of analogous atoms the same. A previous version of *XD* (*XD2006*) could also apply symmetry constraints, which constrained multipole values based on an assigned symmetry of the atom in the molecule, however, this feature is no longer accessible in the program, according to more recent versions of the manual (*XD2015* and *XD2016*).[43]

In an investigation using a benchmark structure (shown in Figure 1.17), which contained a number of molecules in chemically similar environments, Krause *et al.*[62] found that removing symmetry constraints, which were previously included in a multipole refinement, resulted in overfitting of their refinement. Serious overfitting occurred after chemical constraints were also removed from the structural models.

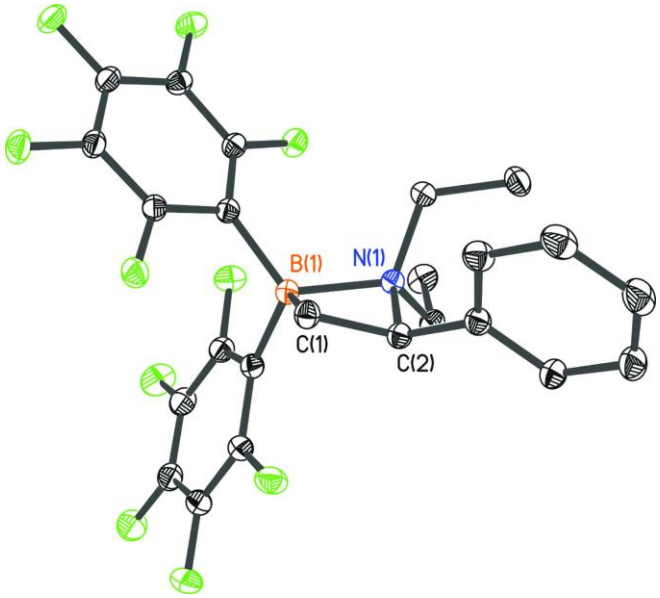


Figure 1.17 Benchmark Structure 1 from Krause et al.,[62] reproduced with permission from IUCrJ.

Cut-offs to the data included in the multipole refinement can be applied by specifying the desired range of reflections in the *xd.mas* file. Data in *XD* can be removed based on the value of the intensity, the significance or the resolution of the reflection. Many charge density studies remove weak data from the refinements based on an arbitrary value chosen by the user. Standard practise, highlighted by the majority of charge density studies in the literature, is the exclusion of reflections with a significance of less than 3. This value is based on a simple statistical significance test, reflections with a significance of below 3 could be reasoned as being not statistically significant from zero.

Watkin describes the debate on the treatment of weak data in his review of structural refinement for small molecules.[84] It has generally been found that weak data have a negligible impact upon the refinement. Despite this, the International Union of Crystallography (IUCr) Subcommittee on Statistical Descriptors recommends the inclusion of such reflections.[83] However, this imperceptible effect on the final structural model is often used as a reason to discard these measurements, although this was not mentioned in relation to charge density data.

In their analysis of α-oxalic acid dihydrate, Kamiński *et al.*,[47] whilst acknowledging that the removal of such values can induce bias into a refinement, chose to omit reflections with an I/σ_I less than 3. They found that the multipole parameters from refinements that included these data were statistically the same as when they were excluded. As Watkin characterises weak reflections into two types, *"those that contain useful information and those that do not"*, Kamiński *et al.* propose that for charge density studies, reflections which could contain information about valence densities (from low $\sin\theta/\lambda$ data) should be included while those from higher $\sin\theta/\lambda$ are up to the discretion of the individual.

Nonetheless, removal of data from a refinement can bias the residual densities as detailed by Henn and Meindl.[42] An I/σ_I cut-off leads to correlations between residuals and their measurement statistics (e.g. I, $\sin\theta/\lambda$), which should not occur if the residual distribution is random (according to conditional probability theory). Schwarzenbach *et al.*[83] state that omitting reflections based on resolution should not introduce bias. However, as high resolution data are critically important for the determination of accurate position and unit cell parameters, it is concluded that the retention of weak data is desirable for charge density studies.

1.5 Reproducibility of charge density models

In 1984, the IUCr's commission on Charge, Spin and Momentum Densities undertook a study to investigate the reproducibility of charge density refinements measured in different laboratories.[88] Oxalic acid dihydride was chosen as the standard sample as it was possible to grow high quality single crystals of the sample simply and reproducibly. The study was undertaken by a number of different research groups; participants grew their own single crystals, measuring and processing the resulting data independently. In total, four X-ray, five neutron, and three theoretical datasets were used for analysis. The authors commented on the small size of the sample but believed that the comparison, although limited, as the first of its kind was still valid for providing an initial estimate into the reproducibility of the technique, especially as there may be differences in the models due to the use of different crystals, instruments and data processing methods.

As the data were processed according to the *"best practise"* of each laboratory, the refinement procedures varied between datasets; for the X-ray data alone, refinements were performed against either F or F^2, with different cut-offs for weak reflections ($I/F < 2/3\ \sigma_I$), along with different equations for extinction corrections and scattering factors. From this

study, it is not apparent if the different processing had any effect on the disparity of the finer detail of the electron density maps or if that solely lay with quality of data measured. Unfortunately, no residual density maps were included, which could indicate the presence of statistical errors within an experiment, so analysis of data quality beyond that which was stated, that the researchers deemed the studies to be of satisfactory quality, is not possible. Two further X-ray data sets collected for the project were excluded from the study for not being high enough quality, implying some level of selectivity of datasets was involved.

The report mainly focused on comparing atom parameters and static deformation electron density maps. Promising results stemmed from the reproducibility of positional parameters (average positional parameter discrepancy between X-ray studies is 0.001 Å) as well as qualitative reproducibility of density maps (although the level of detail varied by experiment). However, less encouraging was the large differences between ADPs and comparisons of finer details within the density maps – namely the variation of the height and position of density maxima. This led to reporters warning against *"over-interpretation"* of the features shown in electron density maps.

A similar investigation was presented at the *'Electron Distribution and Chemical Bonding'* Gordon Research Conference in 2007.[89] The aim was to repeat the previous IUCr comparison using newer detector technology. Diffraction data from single crystals of tetracyanoquinodimethane (TCNQ) were measured on five different experimental setups at 100 (2) K. TCNQ was chosen as a standard for its centrosymmetric structure and the presence of few atoms in the asymmetric unit (ensuring a high data/parameter ratio). Unlike the previous IUCr study, multipolar refinements were implemented using the same procedure for each dataset.

Again, qualitatively there was good agreement of static deformation density distribution across the five studies. However, residual density maps (which were not presented for the IUCr study) showed a range of residual peak heights, as well as a large difference in the presence of "*features"* within the map, indicating the presence of systematic errors within the model and/or data that electron density was poorly modelled in some refinements (Figure 1.18).

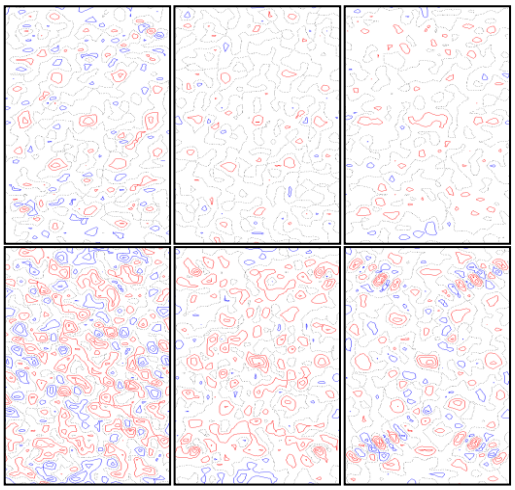


Figure 1.18 Residual density maps from 5 different collections of TCNQ (with the sixth arising from using dataset 5 with different refinement model), reproduced with permission from Probert.[89]

Dipole moments were calculated across half of the TCNQ molecule, to show potential 'knock-on' effects from differences in multipole refinements. The dipole values ranged from 0.84 to 2.67 D, with a standard deviation of 0.7 D (full results shown in Table 1.2). The GooF and R_1 are also reported. If experiment 3 is ignored (which has a GooF of below 1, suggesting possible over-parameterisation of study), there appears to be a linear correlation between increasing GooF with increasing dipole moment. Dipole moments have not been reported in the literature, so it is not possible to compare these results to an accepted figure. An additional study, where data from experiment 5 were reduced and merged in a different manner, produced a dipole moment of 5.36 D. This suggests that both the quality of data collected and the way in which the data were processed can propagate errors through the refinement process which lead to a significant variability in the calculated parameters.

Study	**1**	**2**	**3**	**4**	**5**
Dipole moment (D)	1.34	0.84	2.67	1.58	1.14
R_1	1.89	1.69	1.77	2.86	2.34
GooF	1.76	1.37	0.99	2.39	1.66

Table 1.2 Selected refinement factors for charge density data of TCNQ collected on several different instruments, reproduced with permission from Probert.[89]

Comparing the GooF values from the study to other charge density studies within the literature, it is felt that these values are rather high – for comparison a spherical atom refinement GooF would be expected to be close to 1. Goodness of fit values were not reported in the IUCr study, so it is not possible to compare these values to what were considered good enough by other researchers.

Probert also suggested the need for new descriptors for charge density refinement quality, as using the standards for the spherical atom model did not give enough information about

the quality of data. Much of the work completed by Henn and Meindl has confirmed this, with the additional descriptors proposed discussed previously (Section 1.2).

Another comparative study of α-oxalic acid dihydrate has been undertaken in which charge density studies from 13 separate datasets, collected over several years on three different diffractometers, were analysed.[47] A number of different properties emerging from multipolar refinements were investigated for reproducibility, including those directly resulting from the structural model (unit cell, multipole and geometrical parameters) and descriptors calculated using the model (dipole and quadrupole moments, electrostatic energies).

It was found that many parameters, including unit cell dimensions, bond lengths and ADPs were not statistically different from previous studies of oxalic acid and results within the analysis.[65,88] Unit cell parameters were found to be normally distributed. When scaled raw intensities were analysed, the correlations between datasets were almost linear. A comparison of standard uncertainties of the measured intensities yielded greater variation. The dynamic deformation densities of the models were also found to be reproducible.

Large standard error values for calculated dipole moments, electrostatic potential distributions and electrostatic energies lead researchers to suggest that these values be used more qualitatively, rather than as quantitative properties that could be compared exactly between studies. The authors suggest additional analysis of reproducibility using different data processing methods should take place, along with investigations on a wider range of instruments and detector technologies, for a fuller picture of reproducibility across the field.

The reproducibility of multipole models is also an important consideration in relation to the transferable aspherical atom model (TAAM). This is another method of creating a model of the electron density of a molecule including the aspherical valence density. Unlike the multipole model, the aspherical densities for each atom are imported from a databank of multipole populations for different atoms. The foundation of this technique is that the multipole populations of atoms in chemically equivalent environments are approximately the same, hence the populations are transferable.[90,91]

Brock *et al.*[91] suggested that as constraining the values of parameters in protein crystallography or bond lengths and geometries within the IAM was deemed acceptable, then it follows that constraining the multipole parameters in a refinement to *"their most*

likely values" should also be possible. Multipolar parameters from a study of perylene were transferred to a number of refinements of two similar molecules – naphthalene and anthracene – for which standard multipolar refinements were not possible, to produce models of the electron density. By applying the multipolar parameters to the refinements, improvements in R factors and ADPs for the datasets were observed. The authors posed that this technique could be applicable to other molecules. Jelsch *et al.* created a database of multipolar parameters from charge density studies of peptide structures and applied this in a refinement of a standard collection of an octapeptide.[92] Like in Brock *et al.*, lower values for the R factors of the refinement – along with improvements to the ADPs – were detected. Jelsch *et al.* later compared an experimental multipole refinement of the protein Crambin to a model derived from parameters from their database and observed a degree of agreement between the two refinements.[93]

A number of databanks have been created using different techniques. ELMAM was derived from experimentally measured high resolution structures of proteins.[94,95] ELMAM2 has been released as an extension to ELMAM, to include additional atoms beyond those commonly found in proteins.[96] The University at Buffalo Databank (UBDB) was calculated using multipole refinements of the theoretical electron density of molecules.[90,97] The multipole populations in the Invariom databank are also produced from theoretical models, but the density for each chemical environment of the atoms is modelled separately (instead of extracting from molecules in UBDB database).[98,99]

As previously stated, using the TAAM allows multipole parameters to be included in a refinement, which usually results in an improvement of refinement statistics – such as the R factor and residuals.[92] Jelsch *et al.* state that as the ADPs in IAMs are biased through partially modelling the aspherical density surrounding the atoms, the use of TAAM in structures unsuitable for multipole modelling could produce ADPs which are more physically meaningful. It has also been shown that using aspherical electron densities can result in an improvement in the mean square values calculated in the Hirshfeld bond test, again suggesting a more physically reasonable model.[45] The TAAM has found use in protein structures,[92] where the resolution limit of the data collection can prevent refinement using the multipole model,[100] allowing the calculation of interaction energies from the models.

These populations have also been used as starting points for multipole refinements.[47] Dittrich *et al.* use the Invariom databank to show that using Invariom derived multipole

parameters for ciprofloxacin hydrochloride 1.4-hydrate, with 0.77 $Å^{-1}$ resolution at 120 K (a *"normal resolution"* dataset) was comparable to a multipole refinement of the same compound measured to ultra-high 1.16 $Å^{-1}$ resolution at 100 K.[101] Dittrich *et al.* conclude that it is possible to use the TAAM with *"normal resolution"* data to determine an accurate model of the electron density as long as *"thermal motion and electron density can be deconvoluted".* Sanjuan-Szklarz *et al*., in a comparison of TAAM and the spherical atom model, advocate for use of TAAM over IAM for low resolution small molecule data to gain better quality structural parameters.[102]

Comparative studies of the databases have taken place. Models created using the UBDB and ELMAM databanks and electrostatic properties calculated from them were compared (by the creators of the ELMAM databank) and described as only having qualitative agreement.[103] The authors of the UBDB responded that this should not imply that one databank produced more accurate results than the other until further scrutiny of the ELMAM databank took place.[104] Johnas *et al.* produced a comparison of an experimental multipole refinement with TAAMs created using the UBDB and Invariom databanks;[105] the TAAMs were found to be comparable to the multiple model, with small differences in the multipole parameters.

Bąk *et al.* compared multipolar refinements to TAAM models created using UBDB, ELAMAM, ELMAM2 and Invariom with the data cut to a resolution of 0.7 $Å^{-1}$.[100] Theoretical electron densities were also calculated and compared. Bąk *et al.* report no significant differences of geometries between TAAMs created with the different databanks and only small differences in the ADPs. A comparison of deformation electron densities from the models showed a significant difference between TAAMs, although there was a global similarity in the shapes of the densities, as well as from those created from experimental multipole refinements. Dipole moments calculated from the models deviate from each other with an average of 2 – 3 D, with a difference in direction of 15° - 30° depending on the database used. The authors make no comment on the difference in derived parameters and the validity of the different databanks.

1.6 Investigation of the effect of data processing in charge density refinements

If multipole refinements are reproducible, variation in sensible data processing choices should have no effect on the final refinement. Additionally, atoms in chemically equivalent

positions within the refinement should have statistically the same multipole parameters. Many studies have suggested this is an area of potential interest to investigate.

Within this thesis, Chapter 2 will describe *CAPOW*, a graphical user interface for calculating the optimal weighting scheme parameters for use in charge density refinements, and the effect of a weighting scheme on the final refinement. Chapter 3 details investigations into the effect of the data processing strategy on the multipole model on an experimental dataset. Multipolar analyses undertaken with several small molecule single crystals are described in Chapters 4 – 6. Additional improvement in refinement methodology can occur when utilising software packages that the user is more familiar with; Chapter 7 outlines the conversion of diffraction frames into a different format for increased ease of data processing. Finally, Chapter 8 will describe an early direction for the investigation of the improvement of charge density data processing – a program for the creation of synthetic diffraction frames. A simple modification of this program showed that fraudulent crystallographic data could be obtained. Further discussion is provided of the impact of fraudulent data in crystallography and methods to prevent this type of fraud, as was presented at the European Crystallographic Meeting 29, in Rovinj Croatia in 2015.

Chapter 2. Weighting schemes in multipolar refinements

In crystallography, a weighting scheme is a function applied to the experimental data that is used to calculate a weight value for each reflection. This weight is a corrected measure of the importance or information content a reflection has within the refinement.[83] The weighting of a reflection, *w*, plays a key role in least squares refinement, where the function to be minimised, *S*, includes a weighting term (Equation 2.1).

$$S = \sum w(F_o - F_c)^2 \qquad 2.1$$

The weighting scheme is almost always, wholly or partially, based on the standard uncertainty value for each reflection. This allows the uncertainty of the measurement of the intensity to be taken into account when determining the structural model. When conducting charge density refinements, a statistical weighting scheme (Equation 2.2) is most often applied – consequently the reflections are simply weighted according to the inverse of their standard uncertainty squared (where I is the structure factor, either *F* or F^2).

$$w = \frac{1}{\sigma_I^2} \qquad 2.2$$

The success of this scheme depends upon the standard uncertainty of each measurement being a true and valid value calculated during the data reduction and processing steps. This criterion is often evaluated using a normal probability plot, where the presence of a non-normal distribution of residuals can imply the existence of systematic errors within a given dataset. While least-squares minimisation methods do not depend upon a normal distribution of residuals from the least squares refinement, the refinements are not robust in cases where these residuals are not normally distributed.[106] Furthermore, uncertainties for any parameters derived from the model, such as atom positions or bond lengths, or molecular properties, such as dipole interactions, are calculated with the assumption that the residuals are normally distributed. Therefore, there is an implied expectation that the distribution of the standard uncertainties is normal.

If the source of any systematic errors in the model is the incorrect estimation of the standard uncertainties,[40,61,66] then more advanced weighting schemes can be used in an

effort to acquire more representative values and hopefully a more normal distribution of residuals.[33] Advanced weighting schemes should only be included at the end of the refinement when the structural model is understood to be complete and correct; any known potential sources of error should be investigated and removed before applying a more advanced weighting scheme. The form of the weighting schemes that can be applied, unless direct manipulation of the standard uncertainties occurs, is dependent on what is available within the refinement package being used.

Within *XD2016* and prior versions there is the option to apply a unit, statistical or a *SHELXL* weighting scheme.[107] In the unit weighting scheme, the weighting value for each reflection is set to 1 and has no dependence on any property of the reflection i.e. all reflections contribute equally to the outcome of the least squares refinement.

The *SHELXL* weighting scheme has six variables that can be defined by the user, which are given as *a* – *f* in Equations 2.3, 2.4 and 2.5. For a refinement on F^2:

$$w = \frac{q}{\sigma_{F^2}^2 + (ap)^2 + bp + d + e\sin\theta} \qquad 2.3$$

Where:

$$p = f \times F_o^2 + (1 - f) \times F_c^2 \qquad 2.4$$

and q varies depending on the sign of variable c:

$$q = 1.0 \text{ when } c = 0,$$
$$q = \exp(c\left(\frac{\sin\theta}{\lambda}\right)^2) \text{ when } c > 0, \qquad 2.5$$
$$\text{and } q = 1 - \exp(c\left(\frac{\sin\theta}{\lambda}\right)^2) \text{ when } c > 0.$$

Further weighting schemes exist and can be applied in other software. *MoPro*,[50] a multipole refinement package initially developed for protein and supramolecular structures, contains a variety of different weighting schemes, including the use of the *SHELXL* weighting scheme; it should be noted that this implementation of the scheme only includes the first two user assigned parameters (*a* and *b*) as shown in Equation 2.6. In this case p is $(F_o^2 + 2 \times F_c^2)/3$ (i.e. f = 1/3).

$$w = \frac{1}{\sigma_{F^2}^2 + (ap)^2 + bp} \qquad 2.6$$

Alongside unit and statistical weightings, *MoPro* has the ability to calculate the optimal parameters for resolution-based weighting scheme to achieve a GooF of 1 for a refinement,

as shown in Equation 2.7, where the *a*, *b* and *c* parameters are variables for the weighting scheme and *d* is the resolution of the reflection.

$$w = \frac{1}{(a\,\sigma^2(I_o) + b\,I_o^2)d^{-c}} \qquad 2.7$$

Caution should be exercised if the value of *c* deviates from zero, as this applies a θ dependant correction to the weights. Values for *a*, *b* and *c* can be input by the user or *MoPro* can compute and apply a new weighting scheme optimising the values for all three parameters, *a* and *b*, or *a* alone.[108] The default settings for these parameters are *a* = 1, *b* = *c* = 0, i.e. statistical weights Another scheme that can be applied within *MoPro* is:

$$w = \exp(-B \times s^2) \qquad 2.8$$

where s = $\sin\theta\,/\lambda$ and *B* is a value related to the isotropic parameters.[109]

The frequency of use of the *SHELXL* weighting scheme in IAM refinements has meant this scheme has found use in some published charge density refinements. In native *SHELXL* refinements with a spherical atom model, optimal values for the *a* and *b* parameters are routinely calculated for data refined using the *SHELX* suite, with the other four parameters fixed. The values of *c*, *d* and *e* are set to 0 while *f* = 1/3, a value which has been shown to reduce bias of weighting scheme as opposed to using F_c or F_o alone.[110]

The addition of multipole parameters in a charge density refinement changes the model and thus the calculated structure factors, therefore, the optimal values for the *a* and *b* parameters calculated from the previous spherical atom refinement will necessarily differ from those after multipolar expansion. While *MoPro* can refine optimal weighting parameters for the resolution-based weighting scheme, both *XD* and *MoPro* currently do not provide a mechanism for optimising parameters for a *SHELXL* weighting scheme. Values for the parameters included in this scheme must be manually estimated and input by the user; values are often chosen to display a wGooF of as close to 1 as possible[111] or a set of residuals that demonstrate a normal distribution.[38]

In order to use this scheme effectively and increase confidence in the uncertainties of parameters derived from refinements, a method for calculating the optimal values for a multipole refinement is required. To this end *CAPOW*, a program for the Calculation And Plotting of Optimised Weights, has been developed.[112]

2.1 Requirements

CAPOW is written in *Python 2.7*.[113] The program can calculate the optimised *a* and *b* values for a *SHELXL* scheme using a structure factor file from a completed refinement. A GUI has been created to allow straightforward operation of the program and provide informative output. The current version of *CAPOW* features a window with two different tabs: one for the calculation of the optimal weighting scheme and another for the creation of a normal probability plot of data from a crystallographic refinement to enable visual inspection of the data.

The *CAPOW* GUI requires *Python2.7* and the packages numpy, scipy,[114] matplotlib,[115] and pyqt.[116] Optimised *a* and *b* values are calculated for refinements on F^2, based on F_o^2, F_c^2 and $\sigma_{F^2}^2$ found within a structure factor file; the program requires structure factor files of type *.fcf* (*SHELXL* LIST 4 or 8) or *.fco* (*XD2016*) structure factor file, which contain all observed reflection indices, the calculated and observed structure factors, as well as the associated experimental standard uncertainties. The program can output optimised values for both *a* and *b* to be applied in further refinements, as well as a normal probability plot of the residuals.

Further information about the refinement is required to perform the weighting scheme optimisation within *CAPOW* and must be extracted from two additional files. The *.cif* file is required to provide the number of independent parameters applied to refinement for the calculation of the GooF. The *.ins*/*.mas* file (depending on the format of structure factor file is supplied) provides the weighting scheme which was previously applied to the refinement (for use in the creation of normal probability plots). The wavelength and unit cell parameters can also be extracted to calculate the resolution of the reflections, if the value is not present within the structure factor file. If the required *.cif* and/or an *.ins*/*.mas* files are not present in the same folder (with the same file name) as the structure factor file selected, these must also be identified by the user; a pop-up box appears asking the user to select the correct file.

The source code has been tested using Scientific Linux 7.0 and Windows 8 and has been written with no operating system dependencies. It is offered as open source under the GNU General Public License v3.0 and is available at https://github.com/nu-xtal-tools/capow.

2.2 Calculation of optimum weighting parameters

The optimal weighting calculation is based upon a script from the computational crystallographic toolbox (*cctbx*).[117] The *cctbx* is open source and freely available to download and alter. To enable easier modification and adaptation, the *cctbx* dependent functions were rewritten to remove the requirement of the *cctbx* package for the operation of the program, however, the actual process for the optimisation is largely unchanged.

Initial starting values for *a* and *b* are calculated using Equations 2.8 and 2.9. These equations are taken from the original *cctbx* function and are referenced as being from the *SHELXL* calculation of the weighting scheme parameters.

$$a_{start} = \sqrt{\max\left(0.001, 0.64 * \frac{x}{\max(1 \times 10^{-8}, y)}\right)} \qquad 2.9$$

$$b_{start} = \frac{0.5 * z * a_{start}^2}{N} \qquad 2.10$$

Where:

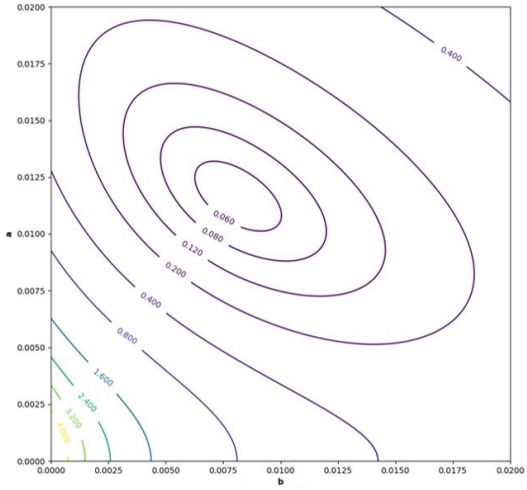
$$x = \sum_{hkl} \left(F_{o,hkl}^2 - F_{c,hkl}^2\right)^2 * \frac{p_{hkl}^2}{\sigma_{hkl}^2}$$

$$y = \sum_{hkl} \left(\frac{p_{hkl}^2}{\sigma_{hkl}^2}\right)^2$$

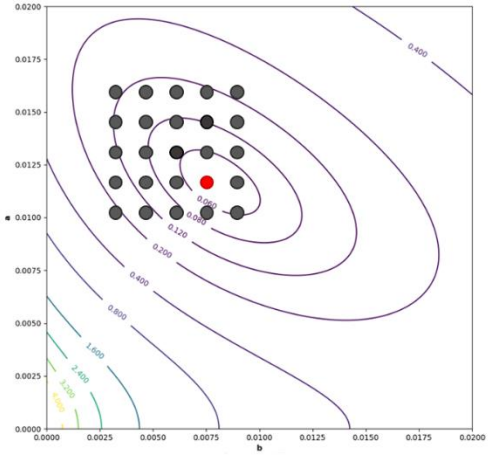
$$z = \sum_{hkl} p_{hkl}$$

The optimal values for the *a* and *b* parameters in the *SHELXL* weighting scheme are determined using a grid search method, a robust algorithm to search for a combination of parameters that gives a minimal value for a function. The process of the grid search method is explained visually in Figure 2.1. Figure 2.1a shows a surface describing the output of a function calculated with different values of parameters (*x* and *y*) within the function. To begin the grid search, the output of the function is evaluated for a series of *x* and *y* values to form a grid of values – initial starting values of *x* and *y* are used and the output of the function with varying *x* and *y* values are calculated, with *x* and *y* increasing in a stepwise fashion, to form the grid. Figure 2.1b shows a grid upon the function surface; a value for the function has been calculated at each grid point and the point with the smallest value is marked in red. The values of *x* and *y* for the smallest output of the function are used to determine a new starting point for a subsequent grid (Figure 2.1c). This process continues with decreasing step size, i.e. how much *x* or *y* increases between subsequent grid points,

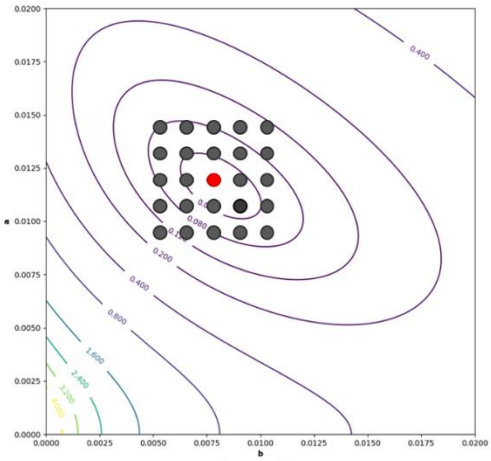
until a stopping condition (usually based on the step size of *x* or *y*) is fulfilled. The values of *x* and *y* which give the smallest output for the function in the final grid are taken to be the optimal values of *x* and *y* (Figure 2.1d).



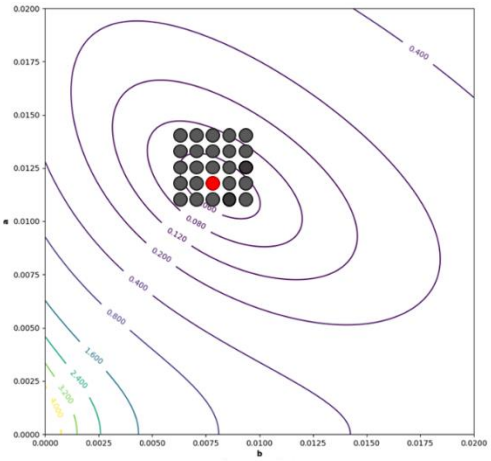
a) A contour plot showing the variation of value X with parameters a and b.



b) A 5x5 grid where the value of X will be tabulated. The red circle indicates the minimum value of X in the grid.



c) The grid is moved based on a new starting point and the values for X recalculated. The step size of the grid can also be adjusted.



d) When a stopping condition is reached, the optimal values of a and b are those where X is at a minimum in the final grid.

Figure 2.1 Diagram to illustrate the process of the grid search method of finding the optimal values of a and b for the weights.

To perform the grid search in *CAPOW*, the reflection data are arranged in order of F_c and divided into 10 equal bins. The function to be minimised is the variance of the wGooF of each of the ten bins. This value is calculated for a 9×9 grid of incrementally increasing *a* and *b* values, beginning with the initial calculated starting values. The initial step sizes are calculated using Equation 2.11. The combination of *a* and *b* that gives the minimum variance within the grid is then used to calculate a new starting point for the next grid – either these values minus 4 times their step sizes or 0, whichever is highest. As a consequence, the values

of the *a* and *b* parameters cannot be negative. The values of a_{step} and b_{step} are reduced provided that the minimum variance did not occur at the edge of the grid.

$$a_{step} = 0.2a_{start} \text{ and } b_{step} = 0.4b_{start} \quad 2.11$$

The process is repeated until the stopping condition is fulfilled. In *SHELXL* the stopping parameters are $a_{step} < 1\text{x}10^{-4}$ or $b_{step} < 5\text{x}10^{-3}$. The user interface in *CAPOW* has been customised to allow a choice of both stopping condition and initial starting values for *a* and *b*. The values of *a* and *b* that give the minimum variance from the final grid when a stopping condition is hit are taken as the optimal parameters to be used in the weighting scheme. The maximum value that can be assigned for *a* is 0.2, if a value above this is returned from the grid search the program reduces *a* to 0.2 and sets *b* value to 0.

Weights calculated using *CAPOW* have been compared to those calculated in *Olex2*,[14] a spherical atom refinement program, which utilises the *cctbx* within the olex2.refine function to calculate weighting parameters. A high degree of consistency between the computed values for each of the routines was found with no significant deviations.

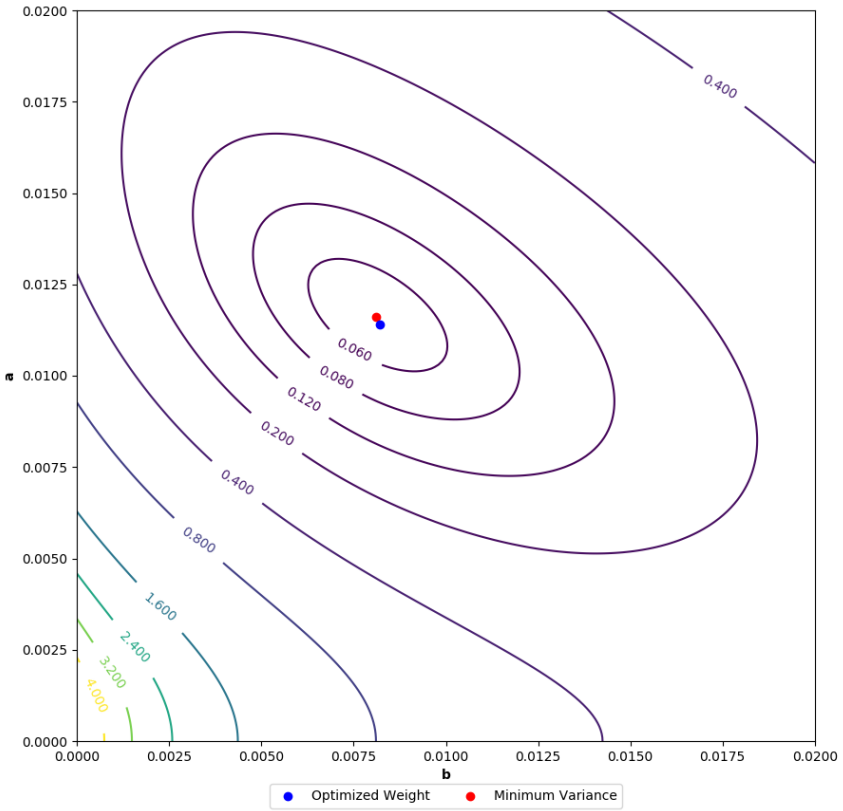


Figure 2.2 A contour plot of the variance of a SHELXL weighted multipole refinement calculated for a range of a and b values. The optimised weight determined by CAPOW and the actual minimum variance of the grid are also displayed.

To assess the applicability of a grid search mechanism for multipolar data, it was important to examine the spread of the variance of the wGooF of binned data – the value to be

minimised in the weighting scheme optimisation – from a charge density refinement with the *SHELXL* weighting scheme applied. If the variance did not tend to a single minimum, it would be possible that the weighting scheme could get stuck in a false minimum or return a non-unique result.

The variance for a grid of varying combinations of *a* and *b* for one multipole dataset is displayed as a contour plot (Figure 2.2). This shows that the variance tends towards a single minimum between 0 and 0.02. The optimal *a* and *b* values calculated by the weighting scheme minimisation (with a_{stop} = 1x10^{-8} and b_{stop} = 5x10^{-7}) correlate well to the position of the minimum of the variance within the grid.

2.3 Normal probability plots

Normal probability plots are created using the *matplotlib* package[115] and are based upon those created in *DRKplot*.[34] The normal probability plots *DRKplo*t produces show the average of sets of binned residuals within the range of -4 to 4 of both the expected and observed residuals. A plot created in *DRKplot* for a multipolar refinement, with statistical weights applied, is shown in Figure 2.3. The blue line included on the graph indicates the expected distribution of points for a normal distribution. The refinement shown displays a non-normal distribution of residuals.

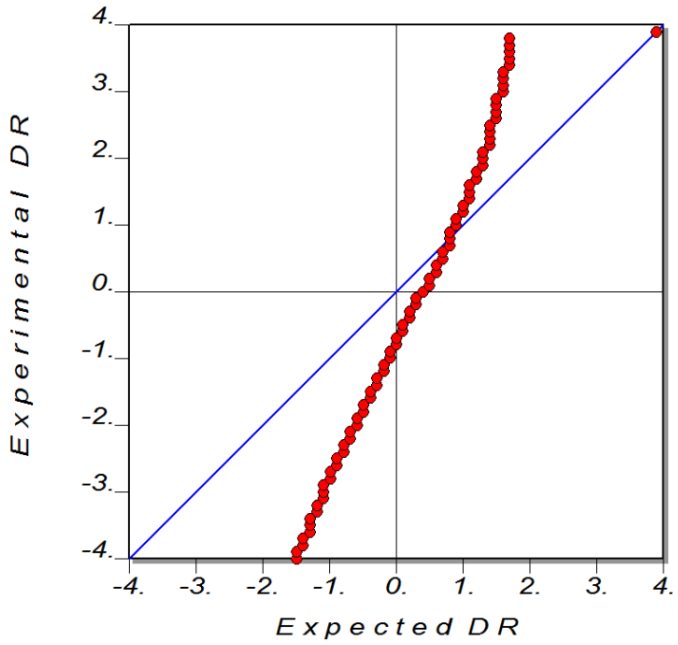


Figure 2.3 Normal probability plot created in DRKplot.

To allow for the identification of potential outliers, additional features were desired in the new implementation. The normal probability plots created in *CAPOW* display every residual value and include the option to edit the range of the axes shown. In reference to their

graphical analysis of standard uncertainties, Henn and Meindl state that using all the data, instead of binning, gives additional details useful in the identification of systematic errors.[40] This notion could also apply to the normal probability plots themselves. While binning the data would not necessarily result in a loss of information in the region observed on the plots, outliers may exist outside of these zones where the individual assessment of reflections may be advantageous.

The *FORTRAN90* code for the calculation of normal probability plots in *DRKplot* was kindly provided by Dr. A Stash. This program was reimplemented in *Python2.7* to provide a comparison for plots containing all values. The residuals used in these plots are calculated using Equation 2.12, where w is the weighting scheme applied.

$$\text{Residual} = \sqrt[2]{w}(F_c^2 - F_o^2) \qquad 2.12$$

For statistical weights, $w = 1/\sigma_I^2$, as shown in Equation 2.13.

$$\text{Residual} = \frac{(F_c^2 - F_o^2)}{\sigma_{F^2}} \qquad 2.13$$

Figure 2.4, equivalent to Figure 2.3, shows that the *Python* implementation recreates the shape and position of the intensities on the plot almost identically and can be used to provide a benchmark for comparison.

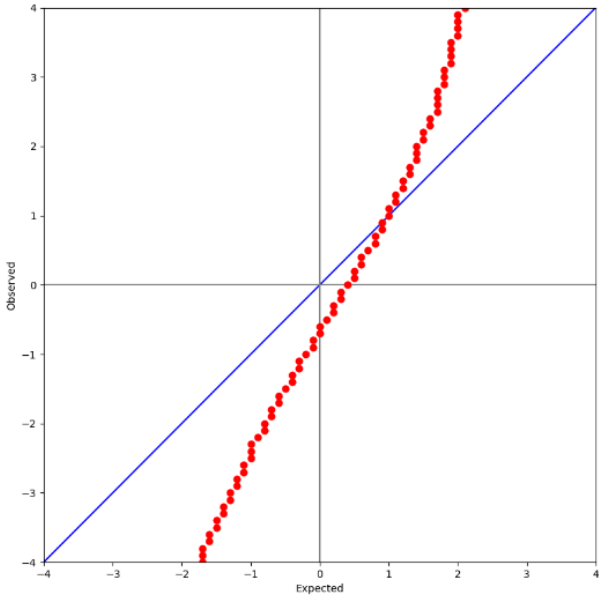
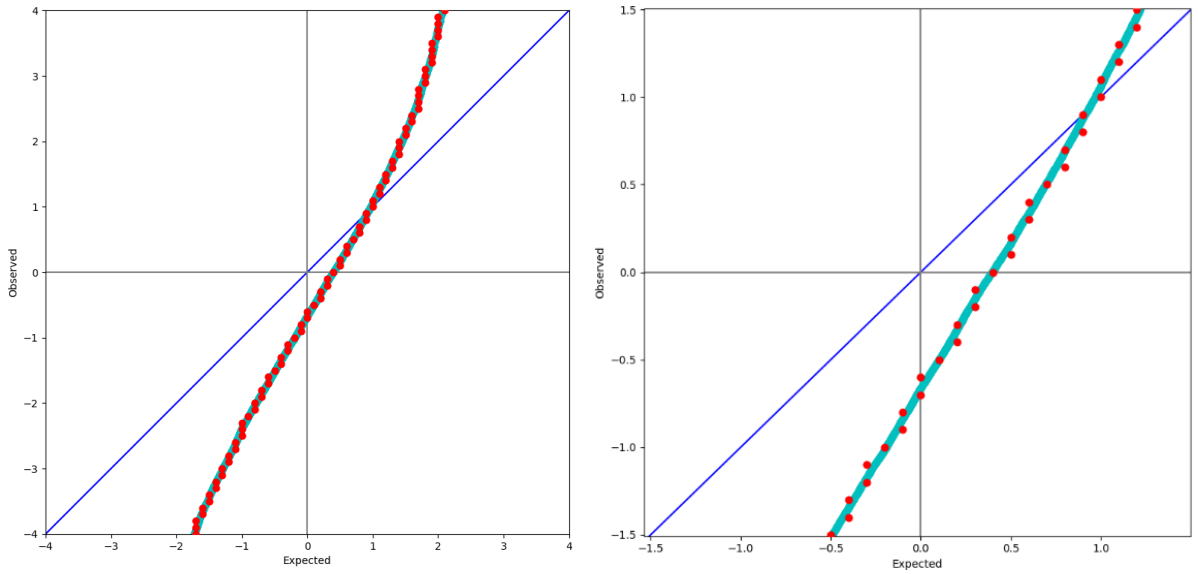


Figure 2.4 Recreated DRKplot-style normal probability plot using Fortran90 translated into Python.

The *DRKplot* benchmark and the normal probability plot of individual values are shown together on the same graph (Figure 2.5). This indicates the distribution of the points is the same between the range shown on the *DRKplot*. Figure 2.5b displays a zoomed in version of Figure 2.5a to make it clearer that the plots demonstrate the same trend. Data showing

individual values has the appearance of a straight line due to the number of points that are present, as the symbols overlap on the graph.



a) DRKplot-style normal probability plot and regular normal probability plot plotted on same graph.

b) Close up of plot a, for closer view of distribution of individual values vs binned values.

Figure 2.5 Comparison of normal probability plot with binned and individual residual values.

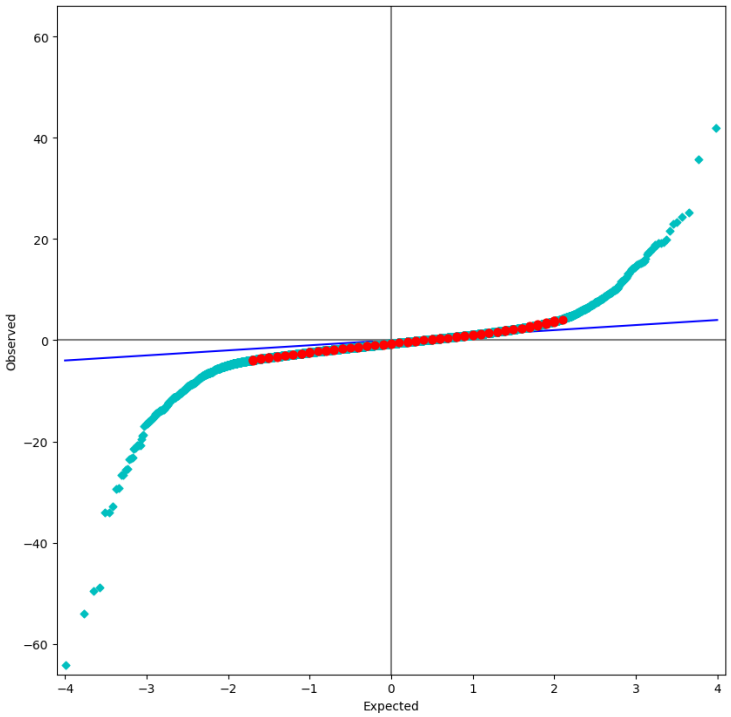


Figure 2.6 DRKplot and normal probability plot showing full range.

Within the range of *DRKplot* little additional information is available by plotting individual points, as the binned values give a good approximation of the distribution of individual points. However, in this particular case and in others like it, outside the range of the plot there are still many points that potentially offer more information. Figure 2.6 is created

using the same data as Figure 2.5, except the full range of residuals is displayed. It can be seen that there are many outliers in this graph, including some observed residuals as large as |60|. The identification of the specific reflections that contribute these extreme residual values could help identify the source of any systematic errors or outliers present. It is clear from Figure 2.6 that there are some systematic errors present for the data within this refinement, including the underestimation of standard uncertainties, that would need to be addressed.

2.4 GUI

The *CAPOW* GUI was designed to make operation of the two scripts more accessible. The ability to apply cut-offs within the GUI prevented the need for users to have to change these values inside the source code and allows rapid comparison of normal probability plots for different weighting schemes.

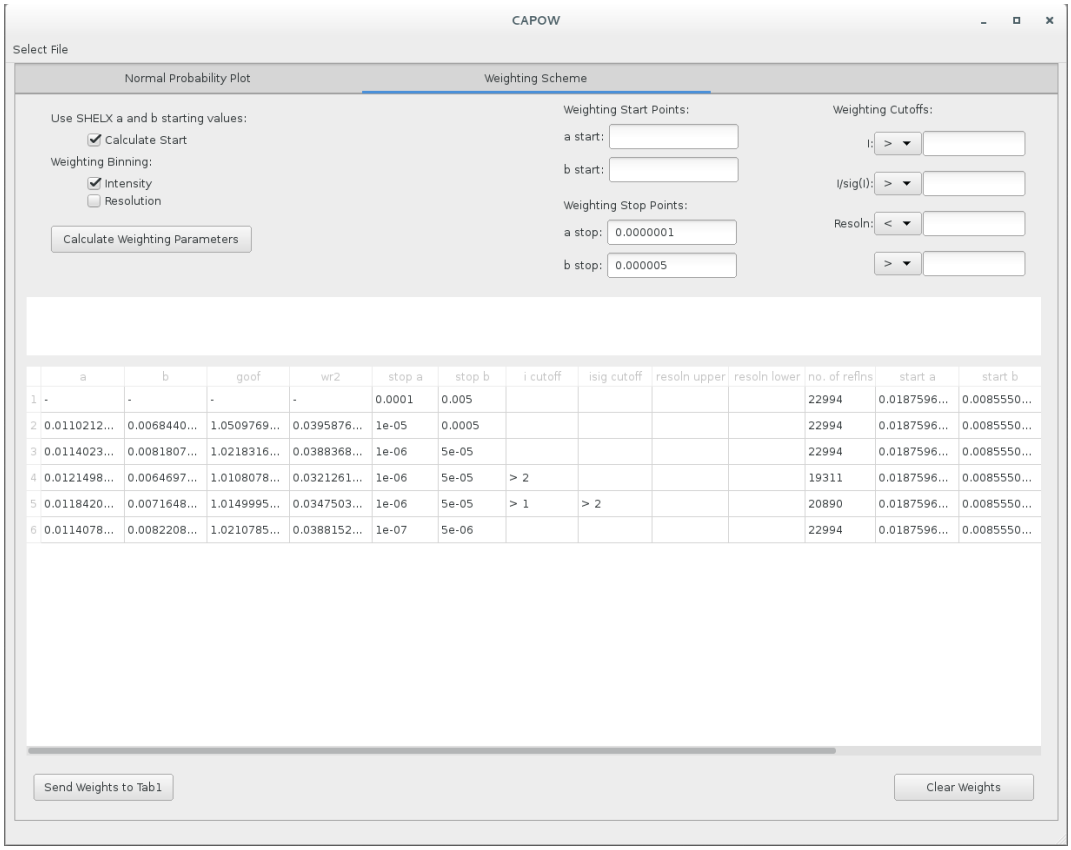


Figure 2.7 Screenshot of the "Weighting Scheme" tab from the CAPOW GUI.

The interface to the weighting scheme calculation is displayed in Figure 2.7. This allows the user to apply various cut-offs to the data from which the weighting scheme should be determined, as well as define the starting and stopping points for the grid search. Starting points can either be input manually or calculated using a *SHELXL* style method (which is the

default method when the "*Calculate Start*" check box is selected). Optimised parameter values for the weighting scheme along with additional information (e.g. the cut-offs applied, the starting points used) are output into a table when calculated. This allows the user to apply various cut-offs and stopping points to calculate optimal *a* and *b* values and compare them.

Selected weights can then be transferred to the normal probability plot tab using the "*Send Weights to Tab 1*" button. The choice for binning the weighting scheme (whether on intensity as in original code, or resolution) was included to investigate any potential differences this change would have on the final weighting values.

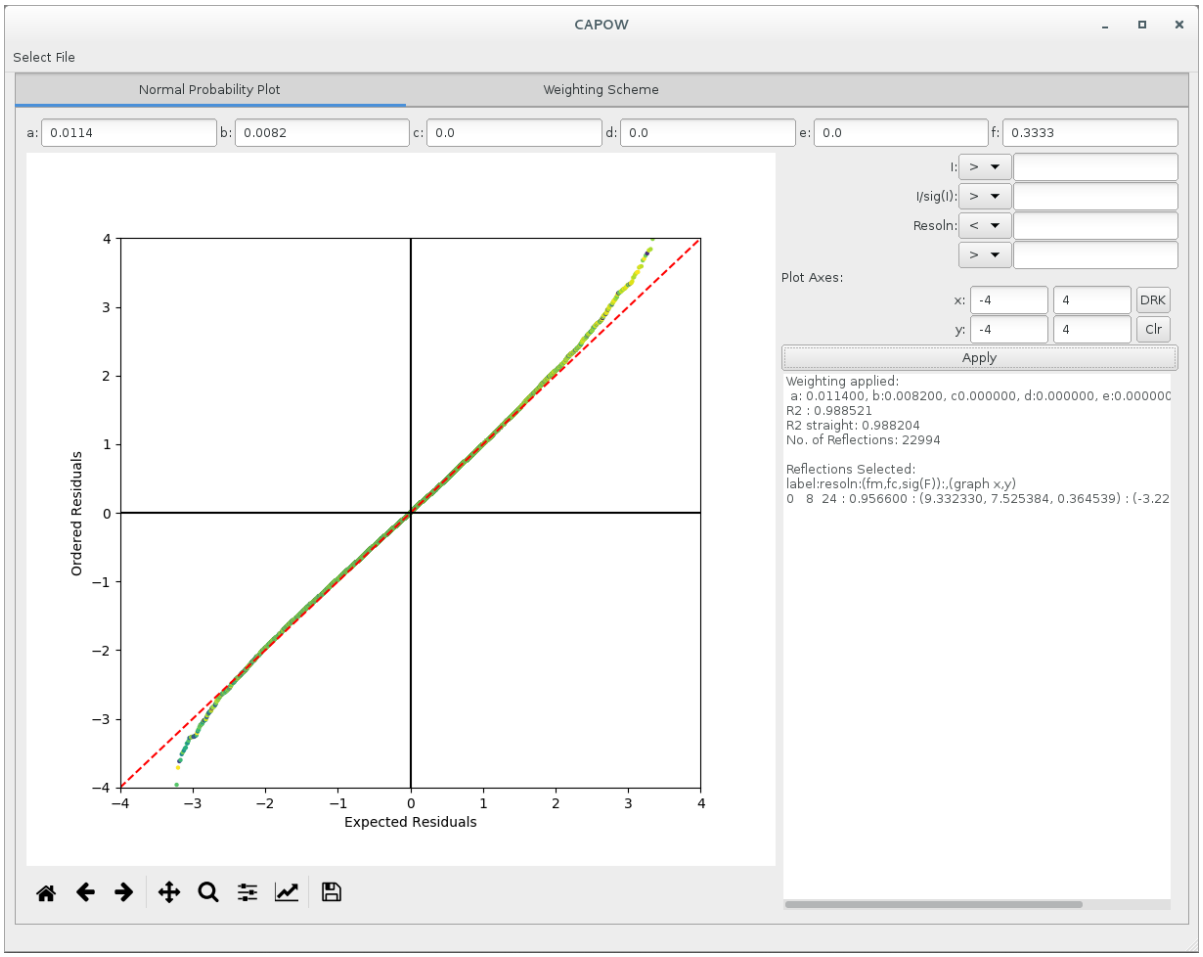


Figure 2.8 Screenshot of the "Normal Probability Plot" tab from the CAPOW GUI.

Within the normal probability plot tab (Figure 2.8) cut-offs to the data based on intensity, intensity over standard uncertainty and resolution can be applied, alongside a weighting scheme which can either be chosen manually or calculated in the "*Weighting Scheme*" tab. An information box allows the user to see the weighting scheme that was applied in the displayed plot. It can also display the reflection indices and structure factor information for any given data point in the plot. The axes of the graph can be edited within the GUI – the

default is to show the whole range of residuals. There is the option to assign the range as requested by the user or to show the graph within the *DRKplot* limits.

The user also has access to all the expected *matplotlib* graph functions through the *matplotlib* toolbar (shown in Figure 2.9), increasing the functionality of the plot. The limits of the graph, alongside the titles of the axes, can be changed by selecting the graph icon on the matplotlib toolbar (second right icon). The toolbar also allows images of the graph to be created and saved for further use.

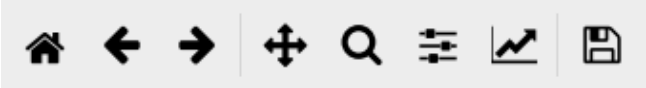


Figure 2.9 Closeup of Matplotlib toolbar.

2.5 Usage of normal probability plot to diagnose errors

The normal probability plot was used to diagnose a problem in the refinement of a charge density dataset of tetracyanoquinodimethane (TCNQ) collected on a CPAD (Photon II), using Ag radiation (λ = 0.56086 Å). A single crystal of TCNQ was slowly cooled to 100 K at a rate of 1 deg/min. The diffraction data were processed with the *APEX3* GUI using the *SHELXL* suite of packages. The data were integrated using *SAINT* and scaled using *SADABS*. The space group was determined using *XPREP*[118] and the structure was solved inside the *OLEX2* GUI using *XT*[119] and refined using *XL*.[107,120] A stepwise multipole refinement was undertaken in *XD2016* against F^2 using statistical weights with the same multipolar data processing method as described for TCNQ in Chapter 6.

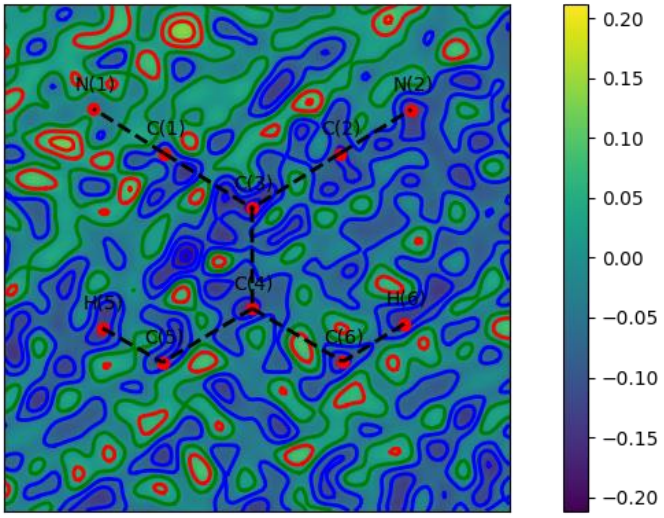


Figure 2.10 2D residual electron density plot for TCNQ.

The final refinement had a large range of residual electron density values -0.212 to 0.139 eÅ^{-3}. When a 2D residual electron density plot was calculated in the plane of the TCNQ molecule, seen in Figure 2.10, a large amount of negative residual density was observed,

suggesting that the model was over-estimating the electron density present in some regions. A normal probability plot was produced which, alongside a non-normal distribution of residuals, showed clear outliers, with many extremely negative residuals (Figure 2.11). This shows that for some reflections F_o is much bigger than the F_c.

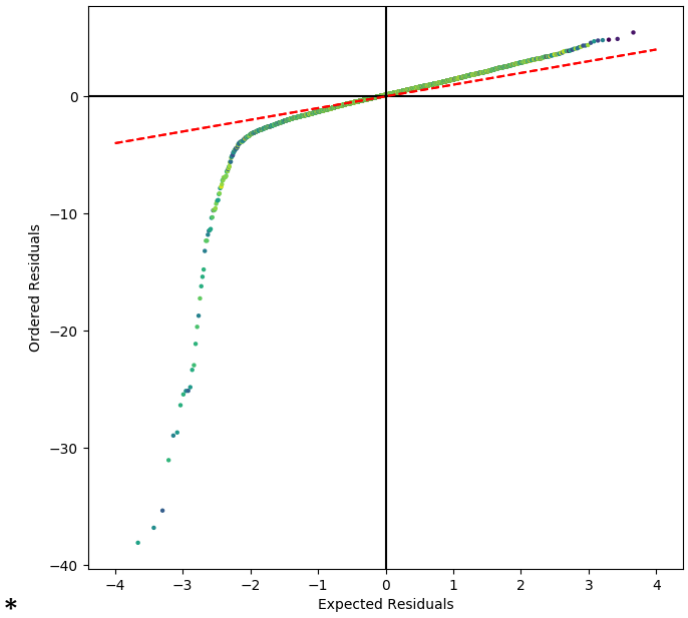


Figure 2.11 Normal probability plot for initial TCNQ refinement.

Using the ability to identify reflections in the normal probability plot in *CAPOW* (by clicking on the specific reflections), the 10 largest outliers were all identified and found to be factors of *4h 4k 4l* (Table 2.1).

hkl	Residual	F_o^2	F_c^2	σ_{F2}
0 8 12	-38.09	92.59	0.09	2.43
0 8 4	-36.69	172.44	102.98	1.89
4 4 4	-35.25	27.02	8.78	0.52
12 4 4	-31.06	34.09	7.05	0.87
4 4 16	-29.32	43.49	20.13	0.80
4 4 12	-28.89	58.87	29.62	0.80
12 4 0	-26. 41	25.62	6.28	0.73
-12 4 8	-25.59	20.22	5.66	0.57
0 0 20	-24.90	21.06	3.05	0.72
-4 4 8	-24.86	233.35	180.25	2.14

Table 2.1 Information about top 10 largest residuals in TCNQ model, which contained systematic errors.

This prompted further investigation; these reflections were probed in the raw experimental data. The problematic reflections were located on the diffraction frames, where it was found that the reflection shape was not of the expected form. Figure 2.12 is an example of this, with the *0 8 12* reflection inside the blue square.

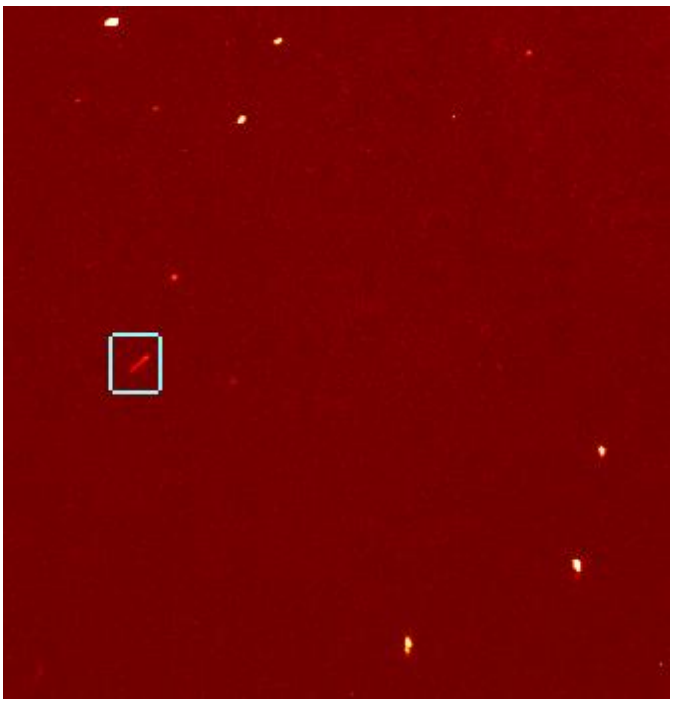


Figure 2.12 Portion of diffraction image showing a measurement of 0 8 12 reflection. The reflection in question is highlighted.

It is possible to view the intensity measured on the diffraction frame in 3D in *APEX3* (Figure 2.13). This provided further evidence that the particular reflections did not have the expected Lorentzian profile.

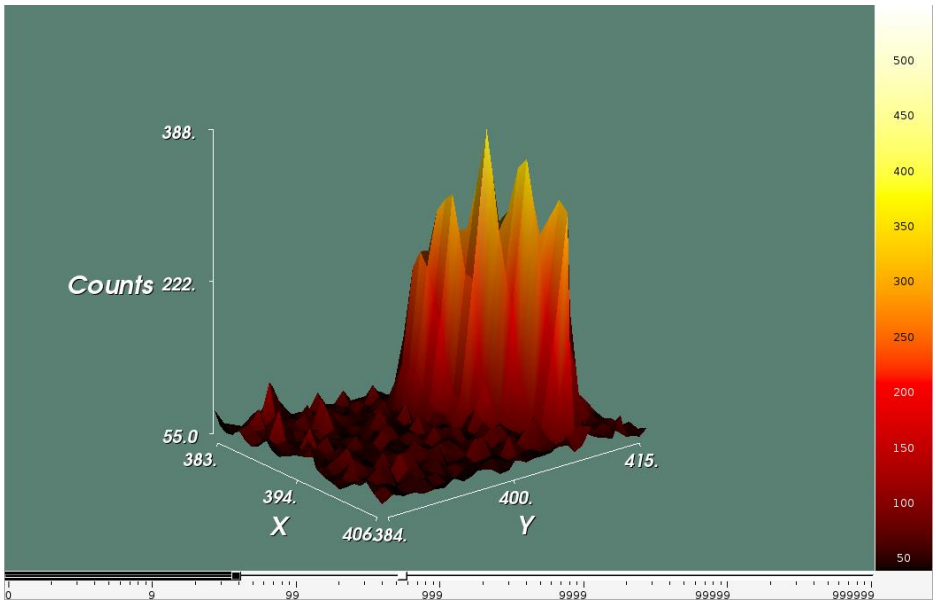


Figure 2.13 3D view of 0 8 12 reflection in contaminated TCNQ data.

A precession image was then created, where all the collected reflections in a specified crystallographic plane are displayed in one image. The precession image in the plane *0kl* (Figure 2.14) contains streaks of X-ray intensity which are centred on reflections with indices *0 8 4l.* These streaks also affect surrounding reflections on the same trajectory as the streak.

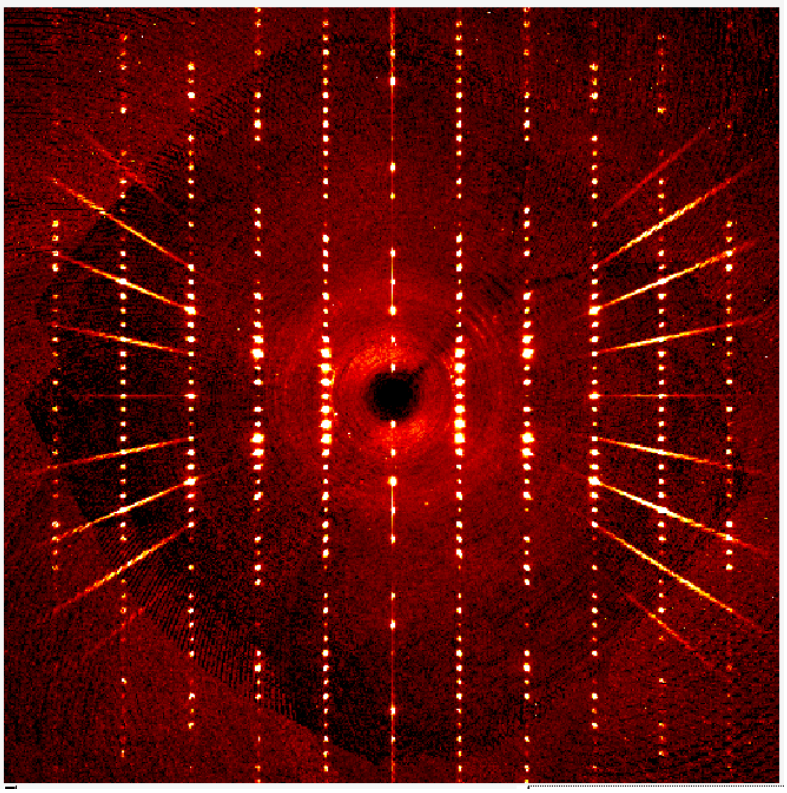


Figure 2.14 0kl precession image calculated for contaminated TCNQ data showing white radiation streaks.

The streaks are caused by contamination of the X-ray beam with low energy photons, so called 'white radiation'. This phenomena has previously been reported for a type of Mo X-ray source which focus the X-ray beam with the same multi-layer optics method as used in the Ag X-ray source.[121,122] Multi-layer optics are thin curved layers of a material which reflect the generated X-rays in order to filter out different wavelengths and focus the final rays to a point. The aim is that only the desired wavelength of X-rays will travel along the correct trajectory to be focused into a beam to be fired at the single crystal. However, this technique does not eliminate photons below a certain wavelength, which are totally reflected by the optics and remain in the beam. These contaminant low energy X-rays cause additional diffraction, which is seen in a precession image as streaks.[122]

For Mo, the contamination was reported with a maximum at *3h 3k 3l* positions in relation to strong reflections,[121,122] however, with Ag the maximum is found at *4h 4k 4l* positions (Figure 2.15). The difference in maxima position is possibly related to the difference in wavelength and hence the longer mirror length and shorter curvature angles required for an Ag source causing a higher energy of photon to be diffracted. The inverse of this reasoning is detailed by Macchi *et al.* as to why Cu microsources would be less affected by this contamination.[121] The contaminant radiation can be filtered from the beam using of a thin piece of

aluminium,[122] removing the low energy contamination at the cost of a small decrease in source intensity.[121]

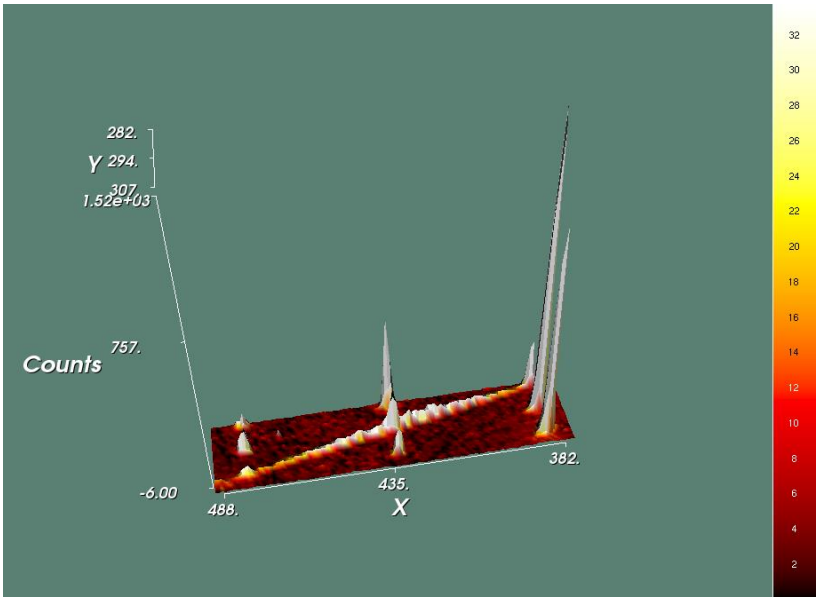


Figure 2.15 3D view of precession image, reflection closest to peak contamination is 0 8 -4.

To assess the thickness of the aluminium filter required to remove the white radiation contamination within the experiment, diffraction data was collected for TCNQ using aluminium filters applied to the Ag source of increasing thickness. Precession images were calculated and assessed until it was found that was a filter of 150µm removed the contamination. The data were then recollected and the shape of the *4h 4k 4l* reflections did not appear to contain any contamination. The multipole refinement completed with the data did not show the identified issues in the residual electron density or normal probability plots. Further details of this refinement are contained in Chapter 6.

Many X-ray detectors cannot discriminate between the energy level of X-ray photons. However, HPADs require an X-ray photon to be above a certain intensity threshold for it to be recorded by the detector. If the low energy contamination was below this threshold, the HPAD would not record these photons; the use of an HPAD instead of a CPAD could prevent this effect being present in the diffraction data.

2.6 Application of weighting scheme to charge density data

The effect of the application of optimised values for *a* and *b* parameters of a *SHELXL* weighting scheme on a multipole model is analysed in detail in Section 3.4.4 and therefore only briefly described within this Section. Figure 2.16 shows the normal probability plot created using statistically weighted data from a multipole refinement. The dashed red line

within the plot indicates the anticipated trend for a normal distribution of residuals. As previously stated, if the residuals were normally distributed a gradient of 1 and an intercept of 0 would be expected, however, a visual inspection of the graph highlights that these data do not demonstrate a normal distribution. The deviation of the residuals from a normal distribution can be assessed by computing a line of regression of the data. In this instance, only reflections between the lower and upper quartiles were used in order to remove bias from any large residual outliers. For the statistically weighted data this gives an equation of y = 1.3985x + 0.02406, further illustrating the difference from an expected normal distribution.

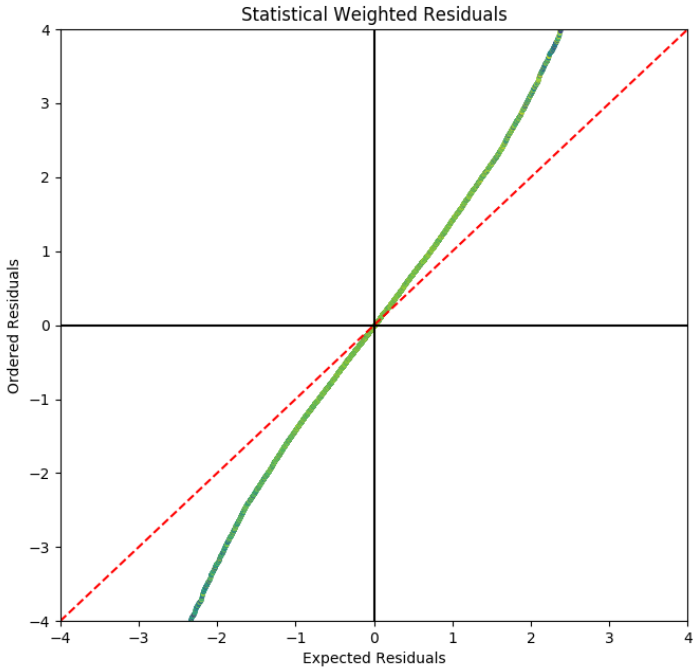


Figure 2.16 Normal probability plot created from a statistically weighted multipole refinement.

To apply the *SHELXL* weighting scheme, optimal parameters determined by *CAPOW* are input into the refinement master file. Several iterations of the least-squares refinement were undertaken until convergence was achieved. Optimal *a* and *b* parameters for the newly converged model are calculated and then applied. The process of calculation and refinement is repeated until convergence of the values of the *a* and *b* parameters occurs.

When using data from a multipole refinement, the stopping points originally applied in the *cctbx* function (a_{stop} = 0.0001 and b_{stop} = 0.005) did not result in convergence of the grid search method. The values calculated for the initial starting point for *a* and *b* (Equations 2.8 and 2.9) lead to the calculation of initial step sizes for the grid search (Equation 2.11) that are smaller than the step size of the stopping conditions (thus preventing the grid search from running). This was a common occurrence observed for the charge density datasets

within this thesis and therefore to counteract this, it is recommended that stopping points of at least $a_{stop} < 0.00001$ and $b_{stop} < 0.0005$ are used. The stopping conditions for the weighting scheme optimisation can be set within the *CAPOW* GUI. The problem could also be overcome by changing initial starting points for *a* and *b*. The weighting applied in this refinement was calculated with stopping points $a_{stop} = 1\times10^{-7}$ and $b_{stop} = 1\times10^{-6}$. Rounds of least squares refinement with optimised *a* and *b* parameters applied in the weighting scheme and re-optimisation of *a* and *b* were completed until the values of *a* and *b* converged to 4 decimal places.

The normal probability plot from the converged *SHELXL* weighted refinement (Figure 2.17) is much closer to that which would be expected from a normal distribution of residuals. The line of regression of data between the lower and upper quartiles has an equation of y = 0.9707x + 0.00006, i.e. a gradient much closer to the ideal value of 1, with an intercept of close to zero.

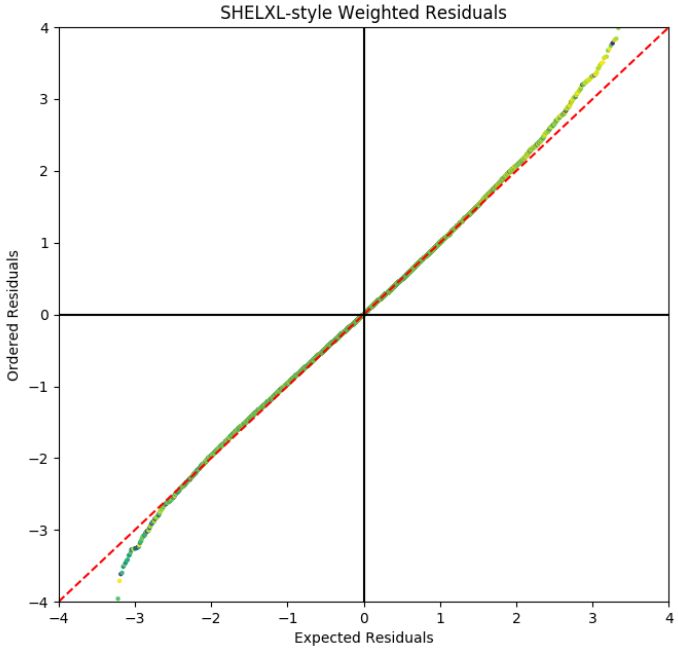


Figure 2.17 Normal probability plot for the same dataset as in Figure 2.16 using an optimised SHELXL weighting scheme as calculated with CAPOW (a = 0.0114 and b = 0.0082).

Key refinement statistics from both the statistical and the *SHELXL* weighting are detailed in Table 2.2. A reduction in wGooF with the *SHELXL* weighting is found, giving a value much closer to 1, along with an increase in the value of wR_2. For the data set used in the above refinements, there was no significant statistical difference observed for atom positions, bond lengths, bond angles or multipolar populations between the statistical and *SHELXL* weighted refinements. However, the multipole parameters calculated from the *SHELXL*

weighted refinement have larger standard uncertainties than the statistically weighted refinement. This causes some populations to become less statistically significant and if these errors are propagated through to other parameters a truer reflection of their accuracy and therefore dependability may be available. Further analysis of the effect of applying a *SHELXL* weighting scheme to a charge density refinement is detailed in Section 3.4.4.

Weighting scheme	Statistical	*SHELXL*
R_1	0.0180	0.0180
wR_2	0.0199	0.0289
GooF	1.6143	1.6679
wGooF	1.6143	1.0118
Residual peak* ($e^-Å^{-3}$)	0.141	0.132
Residual hole* ($e^-Å^{-3}$)	-0.107	-0.094

*Table 2.2 Table of refinement statistics from statistical and SHELXL weighted refinement. *Residual peak and hole calculated using XDFOUR in XD2016 for 2D residual electron density plot.*

2.7 Limitations and future work

CAPOW allows the calculation of optimised values for the *a* and *b* parameters of a *SHELXL* weighting scheme for use with the charge density refinement software *XD*, which does not currently provide a method to optimise these values. There is additional compatibility for programs which can output structure factor files in *.fcf* format. The ability to adjust the stopping conditions of the weighting scheme optimisation allows the process to be completed for the example listed above and other charge density refinements within this thesis, which would not otherwise be able to take place. Normal probability plots created using all variables instead of binned variables allow the analysis of outliers in residual values and can be used to diagnose errors which would not be highlighted using normal probability plots produced with *DRKplot*, the existing software for the creation of normal probability plots.

However, one limitation of this software is that the weighting optimisation has also not been implemented for refinements performed on *F*. This is because the weight applied in *XD2016* is not the same as implemented in *SHELXL* (Equation 2.14).

$$w_F = \left(F_o \times \sqrt[2]{w_{F^2}} + \alpha\right)^2 \qquad 2.14$$

Where w_{F^2} is the *SHELXL* weighing scheme and:

$$\alpha = 0.0 \text{ if } \beta < 0.0;\ \alpha = \sqrt{\beta} \text{ if } \beta > 0.0$$

$$\beta = (F_o^2 \times w_{F^2}) + \sqrt{w_{F^2}}$$

Additionally, for increased practicality it would be useful to create an automated script that ran rounds of refinement, calculated the correct *a* and *b* parameters and applied them instead of having to do multiple stages by hand. *CAPOW* could also be extended to work with *MoPro* input files – the program currently only works with the structure factor files output from *XD (.fco)* or *SHELXL* (*.fcf*). Moreover, the *SHELXL* refinement has three other parameters for which values could also be optimised (*c*, *d* and *e*). Further work could be conducted to analyse the impact of these parameters and to expand the optimisation routines to calculate these additional parameters.

There is the question of whether the *SHELXL* weighting scheme is appropriate to apply in charge density studies. In the *XD* manual,[43] it mentions that the *SHELXL* weighting scheme was designed for a spherical atom model refinement and *"thus may not be adequate for multipole* refinement".[43] Dunitz, in a discussion of weights for least-squares refinements, states an incorrect weighting scheme would only introduce bias onto "*estimates of variances on the parameters*".[123] Therefore, the model should not be affected by applying the weighting scheme, only the standard uncertainties used and then the errors on any calculations calculated from that model. The caveat to this is a situation when applying a weighting scheme too early in the refinement causes a false minimum to be reached. Weighting schemes other than statistical weights are used within the community.[38,124] Zhurov *et al.*[38] suggest that it is reasonable to use a more advanced weighting scheme than that of statistical weights in order to achieve a normal distribution of residuals as the poor determination of standard uncertainties is a known issue.

In this chapter, it has been shown that applying a *SHELXL* weighting scheme can produce a more normal distribution of residuals. However, caution must be exercised as by applying a weighting scheme partially based upon the model we are assuming that the model is correct. Therefore, as a non-normal distribution of residuals can indicate the presence of systematic errors, it is always advisable to use statistical weights initially and eliminate as far as possible any systematic errors before applying a more complex weighting scheme.

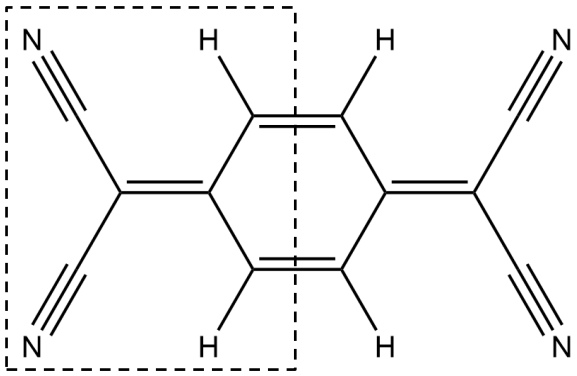
Chapter 3. An analysis of the effect of data processing on the multipole model

The basic approach to data processing for multipole refinement was discussed in Chapter 1. Due to the subtlety of the parameters being refined in a charge density refinement when compared to the IAM, the method of data processing and choice of corrections applied in the refinement become more important. There is no single clear method of processing data for multipolar refinement, with a variety of different parameters and resolution cut-offs applied throughout the field.

There are many different options in terms of which algorithms and corrections to apply, the extent and form of any data cut-off, the choice of parameters to include, as well as which programs to use throughout the data processing and refinement of a multipole model. Invariably some of the choices made will be tailored to a given dataset and the composition of the crystal, such as the level of multipole expansion and specific multipolar parameters included, but others depend only on the preference or 'best practise' of the user. In this chapter, the effect of different combinations of options applied to data processing is investigated with a focus on whether the multipole models are reproducible. This approach has been chosen to highlight the robustness of the multipole models that are refined from diffraction data.

A comprehensive study of this kind, investigating the effect of data processing options on charge density refinements, has not previously been published, although the effects of data collection and instrumentation have been discussed.[65] Kamiński *et al*. undertook a detailed comparison of the multipolar parameters and properties derived from high resolution studies of α-oxalic acid dihydrate, however, these studies were all processed in the same manner.[47] There have also been studies investigating the best processing technique for a specific crystal structure,[125,126] but not with a focus on the reproducibility of data processing. If models are not reproducible i.e. multipole parameters from different refinements are statistically different, then this would lead to a fundamental question about the effect of data processing on the final model and ultimately the validity of any given model.

The choice of sample is very important for the investigation into the reproducibility of multipole models. Small molecules such as oxalic acid and urea have been used widely in multipolar comparison studies,[65,127] because of their simplicity,[128] stability,[88] and the ease of crystal growth. 7,7,8,8-Tetracyanoquinodimethane (TCNQ), has been identified as a potential standard for assessing the quality of charge density determinations by Probert due to the availability of high quality single crystals, the centrosymmetric crystal structure and the small number of atoms in the asymmetric unit of the crystal structure (10 atoms i.e. half of one molecule, see Scheme 3.1).[89]



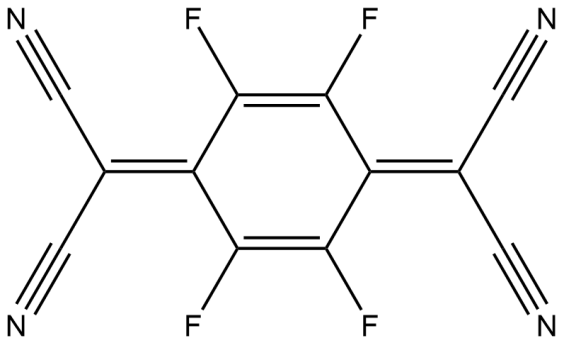
Scheme 3.1 7,7,8,8-tetracyanoquinodimethane (TCNQ) where asymmetric unit is indicated by dashed lines.

However, a drawback in the use of these compounds is that they contain hydrogen atoms. The treatment of hydrogen atoms is non-trivial in X-ray crystallography due to their lack of core electrons, which hampers the precise location of the associated atomic positions.[1] For a covalently bonded hydrogen atom, the centre of electron density pertaining the atom is not typically located around the nucleus, as is the assumption of the spherical atom model, but instead is drawn further away from the nucleus – closer to the atom it is bonding to. As X-ray diffraction patterns arise due to the interaction of electrons with the X-ray beam, a diffraction pattern therefore provides information on the electron distribution within the molecule; hydrogen atom positions assigned from peaks of electron density in the model often give rise to X-H bond lengths that are shorter than expected.[1]

A historical approach to solve this problem has been to use neutron scattering techniques, instead of X-rays, to accurately locate the positions and displacement parameters. Neutrons interact with atomic nuclei instead of electrons, therefore, neutron studies can provide more accurate atomic positions, even for hydrogen atoms.[19] The corresponding neutron experiment must be performed at the same temperature as the X-ray study and scaled appropriately (a non-trivial process). However, few neutron diffraction studies have been carried out relative to X-ray experiments - consequently hydrogen positions must often be

derived from the original X-ray data. There are a number of studies within the literature that suggest different methods for the treatment of hydrogen under these circumstances.[126,129–131] Any method of modelling hydrogen atoms should be applied carefully, as although each atom contributes little to the scattering the total effect can impact on all topological parameters within the model.[129,131]

To circumvent this issue, a compound which did not contain hydrogen atoms was desired. The usage of the fluorinated counterpart of TCNQ, 2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane (F_4TCNQ, Scheme 3.2), as a standard allows the favourable properties of TCNQ to be exploited with the additional benefit of containing no hydrogen atoms. F_4TCNQ crystallises in a centrosymmetric space group and produces single crystals from a variety of solvents which are suitable for multipolar refinement. In addition, F_4TCNQ is a molecule of particular interest to materials scientists. It has a high electron affinity[132] and stable anion form, making it suitable for use as a p-type dopant for a range of semiconductors.[132–134] F_4TCNQ has also found use in charge transfer complexes as an electron acceptor.[135,136]



Scheme 3.2 2,3,5,6-tetrafluoro-7,7,8,8-tetracyanoquinodimethane (F_4TCNQ).

To determine the effect of data processing on a multipole model, diffraction data from a single collection of F_4TCNQ will be processed in a range of different ways with varying options in the integration, scaling, and multipolar refinement stages. The final multipolar refinements will primarily be compared to a reference processing of the data. Refinement statistics and properties calculated from the final multipole models will be used to discern the reproducibility of the multipole parameters when the data processing is varied.

3.1 Data collection

The F_4TCNQ used in this investigation was purchased from Apollo Scientific as a solid with 97% purity and used without further purification. The raw material was a bright yellow powder, unsuitable for single crystal X-ray diffraction experiments. Therefore, crystallisation

experiments were performed to grow high quality single crystals. The most successful of these were achieved via slow cooling of a saturated solution of acetonitrile (purchased from Fischer Scientific, HPLC Gradient grade) over a period of 1 – 2 days. F_4TCNQ crystallised in the orthorhombic space group *Pbca* and was observed to have a regular morphology of yellow block-like crystals.

To minimise as many sources of experimental error as possible, the experimental data collection procedure was chosen carefully. A single crystal of F_4TCNQ (Figure 3.1) was mounted on a *MiTeGen* loop using Fomblin, a low molecular weight perflouropolyether oil. The *MiTeGen* loop is made of an X-ray transparent polymer, chosen to minimise the experimental noise from the mount, while Fomblin is used to reduce the movement of the crystal during the experiment, whether due to gravity or disturbances in the air flow around the crystal due to the cooling equipment. The scattering of X-rays due to the mounting oil adds to the experimental background, however the low molecular weight and viscosity of Fomblin mean this contribution is small.

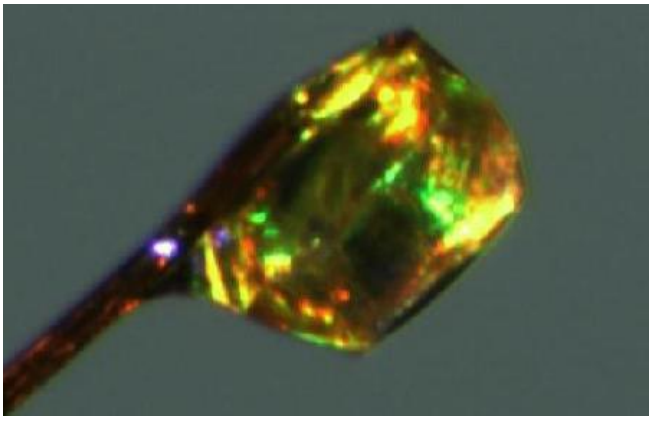


Figure 3.1 F_4TCNQ Crystal as mounted on diffractometer (0.158 x 0.222 x 0.275 mm^3).

The loop was mounted upon a Bruker D8 diffractometer and the crystal was slowly cooled to 100K at a rate of 1K per minute using an Oxford Cryosystems N2 cryostream. X-rays were generated using an Incoatec IμS 3.0 Ag source (λ = 0.56086 Å) with a 150 μm aluminium filter included to remove white radiation contamination before the beam impinged on the sample, a fundamental problem associated with small take-off angle X-ray sources coupled to multilayer optics as described in Chapter 2 (and demonstrated with TCNQ). This problem was present in the F_4TCNQ experiment due to the arrangement of molecules within the crystal, leading to several high intensity reflections that could cause similar scattering.

The diffraction pattern was measured on a Photon II CPAD detector using the shutterless operation mode, which removes the source of experimental error associated with the shutter opening and closing many times during the experiment. The sample to detector distance was set to 65mm to optimise the trade-off between the separation of the reflections and the loss of intensity.

Before an initial strategy could be calculated, a small number of diffraction frames were measured with an exposure time of 10 seconds with the detector at two different 2θ positions to allow the calculation of the orientation matrix of the crystal and measurement of initial values for the unit cell parameters. This information is required to calculate a strategy, or list of experimental runs to be collected, in order to achieve a dataset with the desired completeness and multiplicity.

A strategy for the collection of data with a high degree of multiplicity to a maximum resolution of 0.45 Å was calculated within the *APEX3* GUI. In an attempt to achieve a high signal to noise ratio, individual frames were collected for up to 40 seconds. Long collections can cause the 'overloading' of pixels at locations corresponding to strong low angle reflections. The long exposure time for a diffraction image leads to the number of photons hitting the X-ray detector at those points exceeding the full well potential of the pixel. The intensity measurement for this pixel is recorded as a maximum value, however, many more photons could potentially have hit that point than were recorded so the true intensity value of the pixel and hence the reflection is unknown. To remedy this, an additional 'fast scan' was recorded, with a shorter scan length (1 s). This scan is used at a later stage within the data processing procedure to provide intensity measurements for reflections that contain overloaded pixels. The full experimental run list is displayed in Section 10.1.2.

Ice was observed to have formed on the surface of the crystal in the later stages of the collection. This was evident in run 22 where bright rings due to the powder diffraction of ice crystals are visible on the diffraction images (Figure 10.1). Ice can form due to water vapour in the air surrounding the crystal condensing and then freezing upon the crystal surface; this is often an impact of long experimental times, where there is more time for ice to form. Frames measured after the appearance of the ice rings (from run 22, frame 255 onward) were excluded from the final processing of the data and not significantly impact on the multiplicity statistics.

3.2 Data processing for reference multipole model

An initial multipole model was created to act as a standard or reference to compare to further models. To generate the reference multipole model, the data collected were primarily processed using the 'default' options of the chosen diffraction software.

3.2.1 Integration

Diffraction frames were integrated with *APEX3* with the default I/σ_I cut-off for profile fitting/simple summation for intensity measurement, the best-plane background and a resolution cut-off at 0.48 Å. The resolution cut-off was ascertained using precession images and by examining frames to identify the diffraction limit for this experiment. This was confirmed in post processing using a preliminary solution based on a partial dataset by evaluating the resolution shell that indicated a mean I/σ_I that was less than 3 i.e. the intensity values were not statistically different (within a range of $\pm 3\sigma_I$) from 0.

3.2.2 Scaling

The data were scaled and merged using *SADABS*. A numerical absorption correction to reflection intensities was applied using an accurate model of the crystal shape, which was created by identifying the positions of the faces of the chosen crystal. The faces of the crystal are well defined, as illustrated by the white overlay on Figure 3.2. Run 1 was identified as a fast scan within the scaling and as such only information for overloaded reflections was extracted from this run (at a resolution of less than 1.5 Å); 9 reflections were replaced with data scaled from the fast scan.

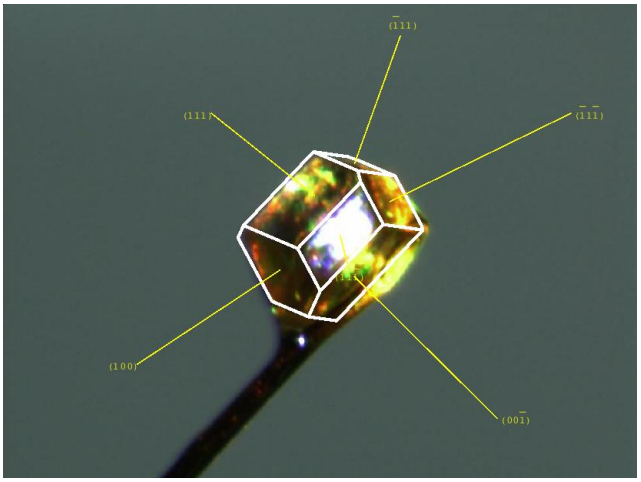


Figure 3.2 F_4TCNQ crystal with faces included. The white area enclosed in one of the surfaces is the reflection of the light illuminating the crystal from a face of the crystal.

The default error model was chosen in *SADABS*, involving the refinement of a global g value with an individual value for the scale factor, K, for each experimental run (Equation 1.28).

Error and parameter models were refined until the value for g converged (g = 0.0133). After the scaling was completed, a reflection file *(xd.hkl)* for use with *XD* was created; reflections with negative intensity, which *SADABS* recommends not to include, and those that should be systematically absent according to the space group were not included in the file.

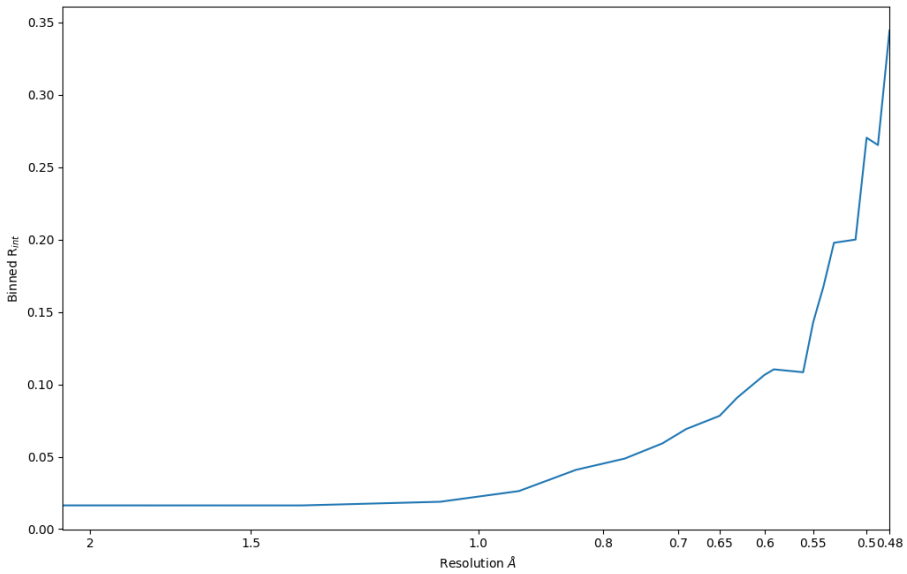


Figure 3.3 A plot of R_{int} against resolution for reference data processing.

To assess how well the individual measurements of the reflections correlate to their averaged values, the R_{int} value for various resolution shells were calculated using the unmerged data. Wolf suggested the creation of this plot as a quality indicator for charge density data (see Section 1.2.1).[27] A plot similar to those created in *XPREP* and *SADABS* was produced using the unmerged version of the *xd.hkl* file of the R_{int} for each resolution shell as a function of resolution (Å) (Figure 3.3) and shows an increase in R_{int} with increasing resolution (Å). The small decrease at ~ 0.55 Å corresponds to an increase in the average I and I/σ_I value for that particular resolution shell; choosing different ranges for the resolution shell either highlights or flattens this feature. This is caused by end of the overlap of detector regions for runs with different 2θ values within the data collection. The time length of the scans in this collection is increased with higher 2θ values of the detector; when the maximum resolution of the frames collected at a lower 2θ is exceeded, the reflections that are scaled together are only present in higher 2θ scans, which may have lower standard uncertainties, resulting in initial greater significance values for the reflections being scaled. The value of R_{int} increases beyond 10% at 0.6 Å which could suggest the high-resolution reflections are weaker and therefore experience increased experimental noise.

3.2.3 Spherical atom model refinement

The unit cell was reduced to the standard setting in *XPREP* with an initial structural model created and refined in the *Olex2* GUI using *XL* and *XT*. After the initial refinement of the IAM had converged there was still an appreciable amount of electron density which had not been accounted for by the model. This can be illustrated using a map of the residual electron density ($F_o - F_c$), as calculated in *Olex2* (Figure 3.4). The residual density is primarily located between atoms, where electron density for covalent bonding may be expected. This density could be accounted for by the multipole model to produce a better representation of the distribution of electron density within the molecule.

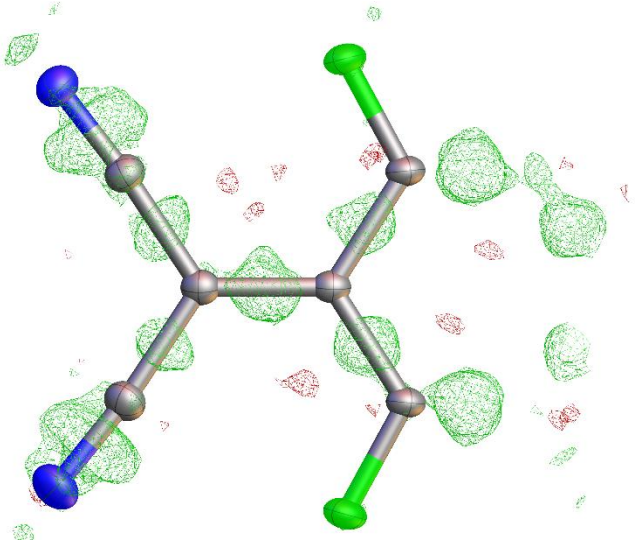


Figure 3.4 Wire map of residual electron density calculated in OLEX2. Wire contour shows ±0.25 eÅ^{-3}.

A summary of the refinement statistics from the completed spherical atom refinement are listed in Table 3.1. This shows that while a good value (< 5%) for the R factor was observed with a GooF of close to one, the range of residual electron density present after a spherical atom refinement was large (0.728 to -0.363 eÅ^{-3}).

Total reflections	6187
Reflections I/ > 3	4828
Sinθ/λ max (Å^{-1})	1.11
R_1 (%)	3.57
wR_2 (%)	11.21
GooF	1.080
Residual density peak (eÅ^{-3})	0.728
Residual density trough (eÅ^{-3})	-0.362
Weighting scheme	*SHELXL* a = 0.0628, b = 0.1144

Table 3.1 Refinement statistic summary for spherical atom model of reference processing of F_4TCNQ data.

The bond lengths as calculated for the IAM within the molecule are listed in Table 3.2, showing the expected alternating quinoidal bond length pattern of the C – C bonds.

Atom 1	Atom 2	Bond length (Å)
F(1)	C(1)	1.3287(4)
F(2)	C(2)	1.3270(4)
N(1)	C(5)	1.1559(6)
N(2)	C(6)	1.1564(6)
C(1)	C(3)	1.4406(5)
C(2)	C(3)	1.4392(5)
C(1)	C(2*)	1.3487(6)
C(3)	C(4)	1.3773(6)
C(5)	C(4)	1.4297(6)
C(6)	C(4)	1.4307(6)

Table 3.2 Bond lengths for reference model from IAM.

3.2.4 Multipole refinement

The refinement of the atomic multipolar expressions was completed using the *XD2016* suite; the *XDINI* program was used to create the required input files (*xd.mas*, *xd.inp*, *xd.hkl*) for the process. The initial input file (*xd.inp*), which contains the values of the various parameters for each atom within the asymmetric unit, was created using the final IAM, refined against only reflections measured at high resolution (< 0.7 Å). This was to enable atom positions to be calculated using reflections that mainly comprise of contributions from core electrons. Atoms within all refinements were identified as shown in Figure 3.5.

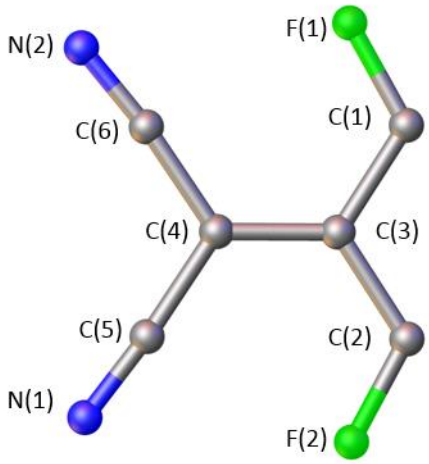


Figure 3.5 Asymmetric unit of F_4TCNQ with atom labels.

The *.hkl* file (*xd.hkl*) is required to be in a different format than the standard files used in the *SHELXL* suite of programs. The *.hkl* file created by *SADABS* for use with *XD2016* was used in place of the file converted by *XDINI*. The initial master file, which contains the information required by the *XD* suite to run the various programs was then edited.

The content of the *XD* master file (*xd.mas*) includes which parameters are to be refined during a given iteration of least squares minimisation and the calculations that are to be

performed, along with various options for data cut-offs that can be selected. The default *XD* databank (CR)[43] was chosen to provide the atomic wavefunctions for the atoms. Least squares minimisation was calculated against F^2 to enable the calculation of the weighting scheme using *CAPOW*. The local coordinate system of the atoms (*X,Y,Z*) axes must also be assigned by providing two vectors (either by the direction of an atom to another atom, or an atom to a dummy atom used to define an alternative coordinate within the unit cell), which are used to calculate a plane within the molecule. The third vector is calculated as perpendicular to this plane to form either a left- or right-handed triplet of vectors, as defined by the user. Using these vectors, an orthonormal vector triplet is formed, which describes the three axes. For the purpose of this study, it was important to ensure that the axes for atoms in similar chemical environments pointed in consistent directions to enable the comparison of the multipole parameters for later processing choices.

Multipolar parameters to be refined in the minimisation process are usually included in the refinement sequentially i.e. only refining the monopole parameters, then including monopoles and dipoles, next monopoles, dipoles and quadrupoles and so on. It is rare that the exact order the parameters were applied in is listed within published literature, including which parameters were refined concurrently. One detailed account was listed in the appendix of a PhD thesis by Wolf.[137] This strategy included iterations of least squares minimisations where different combinations of the final parameter list were refined before parameters were refined altogether, presumably to prevent the correlation between multipole parameters influencing the final refinement. The degree of correlation between parameters, i.e. when the value of one parameter depends highly on that of another parameter, can be assessed in the *XDLSM* output file, where pairs of parameters with a correlation of above a user defined value are reported. Preventing the correlation of parameters should also reduce the likelihood that the refinement will find itself in a local minimum instead of the global minimum of the refinement. A similar strategy to that employed by Wolf was adapted for use with the reference refinement process. A list of the orders of parameters used for this refinement (and others in this chapter) is given in Section 10.1.3 and summarised briefly below.

Least squares minimisations were performed applying a standard cut-off of reflection data based on the significance of the reflection ($I/\sigma_I \leq 3$). Therefore, reflections with a ratio below 3, those with an insignificant intensity with respect to their standard uncertainties, are not

used for the minimisation. An I/σ_I cut-off of this amount is applied by default within *XD*. The general limit or end point to the addition of parameters is the inclusion of monopole and one bond directed dipole for hydrogen atoms, up to and including octupoles for elements in the first row of the periodic table, and up to hexadecapole level beyond this. In this case, even though all atoms are first row elements multipoles up to hexadecapole level were included. Kappa parameters, for the expansion or contraction of the spherical (κ) and aspherical (κ') valence density, were then included and refined. An overall κ' can be determined for the aspherical density or separate values of κ' can be refined for each individual level of multipole, although this is only recommended for transition metal atoms as the refinement is sensitive to the inclusion of κ' – which can be highly correlated to other parameters.[43] As F_4TCNQ only contains elements from the first row of the periodic table, only an overall κ' value was refined for the aspherical valence density. Iterations of the refinement were performed with varying sets of parameters to reduce the impact of correlations between parameters (as in Wolf strategy).[137] The weighting scheme applied was restricted to statistical weights.

To enable a rigorous comparison of the effects of data processing on the final model, it was important to establish a refinement approach common to all datasets. In order to achieve this, a *Python* script was written to automate the multipole refinement process, allowing specified parameters to be refined in a well-defined order. An input file containing the parameters to be included in each iteration was required. The parameters were written into the *xd.mas* and the *XDLSM* process repeated until convergence (defined as the largest shift in parameter as reported by *XDLSM* output being less than 1×10^{-5}). In some cases, convergence with a specific set of parameters could not be achieved, commonly identified when the refinement returns the same maximum shift in the value of a parameter (of a value > 1×10^{-5}) repeatedly at the end of a series of multipole refinements. The program was designed to stop and display an error message highlighting this when the final shift returned was the same for 3 iterations of the refinement. Salient files, such as the model and the output from the refinement are then copied and saved with an alternate name, indicating which iteration of the minimisation it pertains to. These files would otherwise be overwritten by *XDLSM*, therefore are kept so that they can be used for future analysis if required. The output of the *XDLSM* program containing the new values for multipole parameters is *xd.res*. Running *XDLSM* multiple times requires that this file be copied to

overwrite the previous *xd.inp*, to ensure that new parameter values can be used in the next round of refinement. This action is also performed by the program.

3.2.5 Validation of the reference multipole model

Table 3.3 contains a selection of some of the key refinement statistics commonly used as indicators of data quality. As has previously been described, these numbers do not provide a comprehensive understanding of how suitable the model is and how well it describes the distribution of electron density, therefore, a rigorous analysis of the model must also be performed.

Total reflections	4670
Number of reflections used	4029
Sinθ/λ max ($Å^{-1}$)	1.042
R_2 (%)	1.55
R_1 (%)	1.48
wR_2 (%)	1.71
GooF	1.9668

Table 3.3 Refinement statistics for reference data processing.

A number of quality indicators, previously discussed in Chapter 1, have been utilised to assess the final multipole refinement. One primary validation method is the Hirshfeld rigid-bond test, which is passed if the pairs of atoms have a difference in the mean square displacement along the interatomic vector of $\leq 1 \times 10^{-3}$ $Å^2$ (for atoms of carbon and heavier). The values of the mean-square displacement can be calculated by *XDLSM* and are displayed in Table 3.4. As the values range from $\pm 4 \times 10^{-4}$, the rigid bond test criteria is met for all interatomic vectors.

Atom 1	**Atom 2**	**DMSDA (1×10^{-4} $Å^2$)**
F(1)	C(1)	-4
F(2)	C(2)	0
N(1)	C(5)	-3
N(2)	C(6)	-1
C(1)	C(3)	-1
C(2)	C(3)	0
C(3)	C(4)	3
C(5)	C(4)	-2
C(6)	C(4)	1

Table 3.4 Differences of the mean square displacement amplitudes (DMSDA) along interatomic vectors for reference processing.

2D residual electron density plots were calculated in the plane of the molecule. These are constructed using the *XDFOUR* program which calculated a Fourier transform of the $F_o - F_c$,

where F_c is the calculated structure factors for the multipole model. The *XDFOUR* output is a grid of the value of residual electron density at regular points within the defined area. The Windows GUI of *XD* (*winXD*) contains a program (*XDGRAPH*) that can draw contour plots of the output from *XDFOUR*, highlighting the areas of minimum and maximum residual electron density within the plane. However, to track the residual density distribution during the automated processing, a *Python* script was written to create residual electron density plots throughout the course of the refinement. A contour plot similar to that created by *XDGRAPH* is shown in Figure 3.6 for the reference model. A colour scheme was then applied with respect to the residual electron density value at each point in the grid in order to show any features in the plot more clearly.

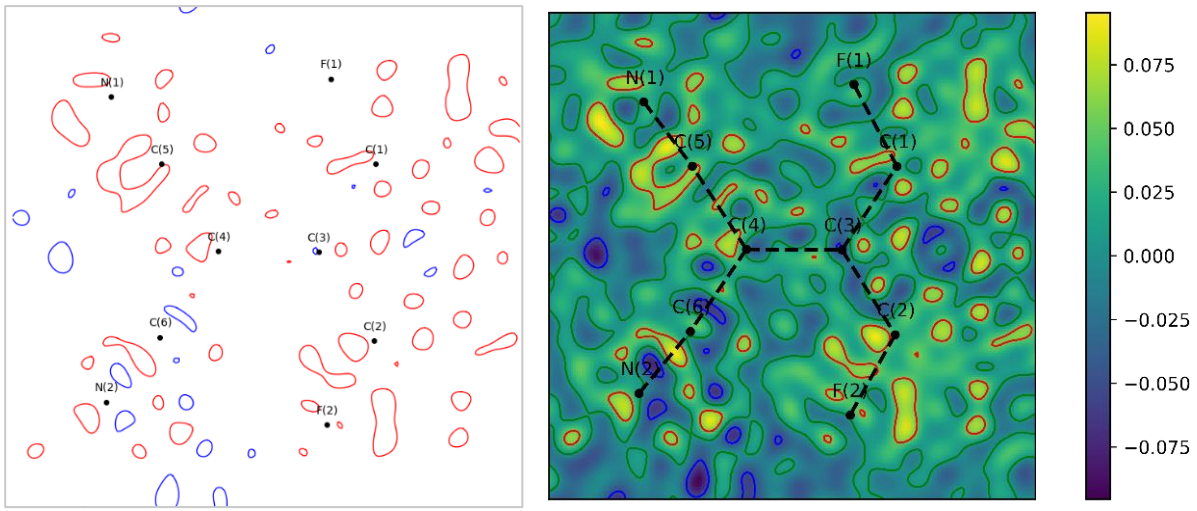


Figure 3.6 Residual electron density plots for the reference data processing showing contours only (left) and with a colour scheme (right). Contours are drawn at ± 0.05Å levels, where red represents positive contours and blue represents negative.

The range of residual density in the plot is -0.093 to 0.096 eÅ^{-3} for the reference model. These values are within the achievable range for light atom molecules theorised by Zhurov *et al.*[38] of ± 0.2 eÅ^{-3} (with $R_1 < 2$). This is also much smaller than the residual density range of -0.362 to 0.728 eÅ^{-3} reported when using the spherical atom model (Table 3.1), illustrating that the multipole model has accounted for more of the electron density present in the structure. Peaks of residual density are distributed about the plane but not solely concentrated around atom positions.

The residual electron density can also be plotted in 3D, which allows a cursory investigation of any features in the density that are present outside of a 2D plane of the molecule. Another *Python* script was created to visualise the contours in 3D by creating surfaces at a desired contour level. Using *XDFOUR,* a 3D grid was calculated for a rectangle surrounding the molecule. This rectangle was 7 x 7 x 3.5 Å^3, chosen to enable examination of the

approximate area relating to the asymmetric unit and prevent the repetition of features due to symmetry equivalent molecules.

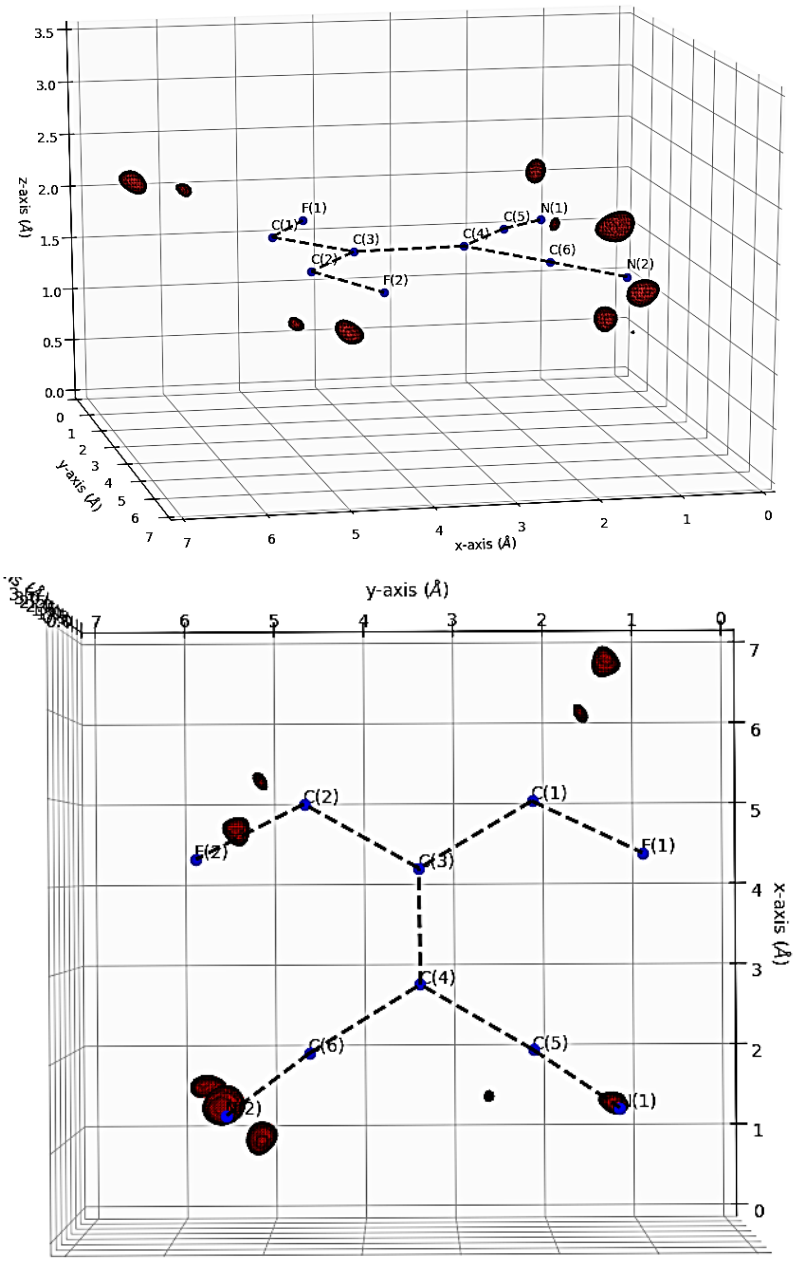


Figure 3.7 Views of the 3D residual density plot for reference data processing with a resolution cut-off of 0.83 Å^{-1}, with contour at ± 0.65 Å^{-1}. Positive contours are shown in red and negative in blue.

A contour plot was created for the multipole model with a resolution cut-off of > 0.83 Å^{-1} as it was felt it was more important to assess the low angle reflections that contain most of the contributions from the valence electrons. This value was chosen as it is approximately the value where the R_{int} rose above 10% (0.6 Å on Figure 3.3). The range of residual values is -0.066 to 0.096 eÅ^{-3} for this grid. The contour plot in Figure 3.7 illustrates surfaces at 0.065 eÅ^{-3}; a few areas of higher residual density occur, but no features within bonds are present.

Including structure factors from higher resolutions in the plot increases the range of residuals and amount of experimental noise in the residual electron density maps.

The plot does show an accumulation of residual density around N(2); attempts were made to reduce this feature. The inclusion of octupolar or hexadecapolar parameters does not affect the appearance of this feature. It can be removed if the κ' parameters are refined individually for only N(2), but the model parameters no longer converge and the value for κ' (O) ~ 1.8, which is a large expansion. This feature is less pronounced when only low angle reflections are used to create a residual electron density map.

Residual density analysis of the 3D data (with no resolution cut-off) was performed using the *jnk2RDAv1.5* program,[37] kindly provided by Dr Julian Henn, and shows deviations from the optimal shape of a parabolic graph (Figure 3.8). There is a slight difference between the two sides of data in the plot, indicating that there is not a symmetrical distribution of values – although visually this appears to be minor.

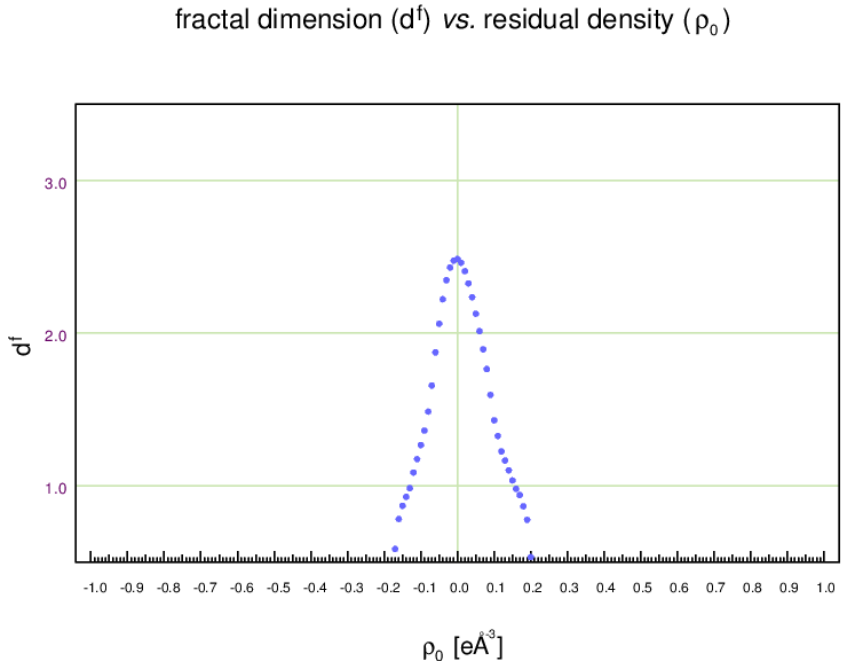


Figure 3.8 Fractal dimensionality plot for 3D grid of residual electron density for reference model (with no sinθ/λ cut-off).

A normal probability plot for the structure factors was created with *CAPOW* (Figure 3.9).[112] The residual values do not follow the expected normal distribution (dashed red line) which indicates the presence of some systematic errors in the data or refinement. The shape of the plot is similar to Figure 1.13, a normal probability plot created using underestimated standard uncertainties,[40] which could imply the underestimation of standard uncertainties within this data as well. The colour of the individual points on the plot indicates the

resolution of the reflection and shows that the largest residual values are predominantly calculated for reflections at low values of sinθ/λ (Å^{-1}).

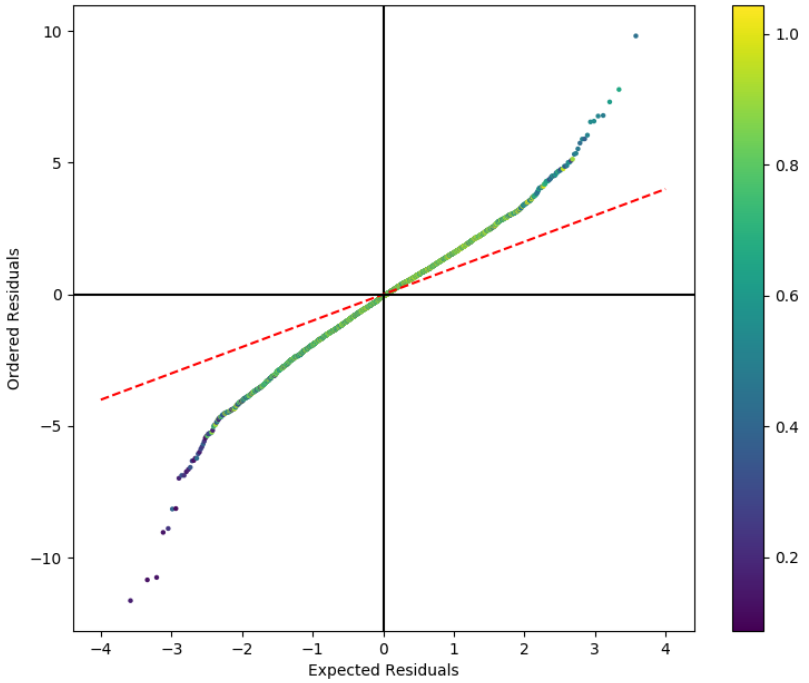


Figure 3.9 Normal probability plot for reference data processing.

The low angle reflections also happen to be the strongest reflections in this dataset and therefore would result in proportionally large differences in the calculated and observed structure factors. This could have a large effect on the least squares minimisation. The residuals calculated for the reflections are negative; for these reflections F_o is larger than F_c (see Equation 2.12). Three possible explanations for this are: reflections are incorrectly measured as too intense, the model does not adequately account for low angle reflections, or the standard uncertainties are not sufficiently large at low angles to account for the potential difference.

The dynamic deformation density, calculated using *XDFOUR*, shows the additional electron density which is accounted for by the multipole model, thereby giving a representation of the modelled aspherical electron density. A visual inspection of a 2D plot calculated in the plane of the molecule (Figure 3.10) shows the expected arrangement of electron density within the molecule. The electron density in the plot ranged from -0.193 to 0.811 eÅ^{-3}; the maximum peak of electron density is larger than that of the residual density present in the spherical model (Table 3.1). This may be due to differences in the form of the spherical atom model structure factors which are subtracted from the multipole model to produce the plot or due to the difference in weights – as the spherical refinement in *XL* was refined with *SHELXL* weights whereas this multipole model is statistically weighted. The brighter areas of

the plot correspond to triple bond cyano groups and the double bonds within the ring, where classically the most electron density would be expected to reside.

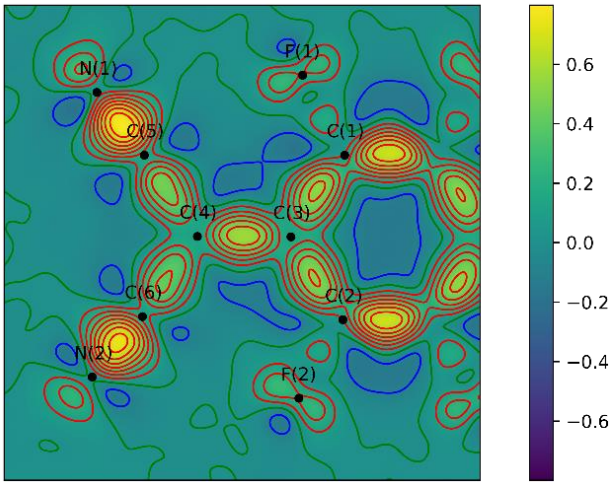


Figure 3.10 Plot of the 2D dynamic deformation electron density for the reference model in the plane of the molecule. Contours are recorded at levels of ± 0.10 eÅ^{-3}, where red indicate positive contours, and blue indicates negative contours. The zero contour is pictured in green.

The static deformation density, Figure 3.11, broadly shows a similar distribution of electron density within the plane of the molecule. The same contour level is used as the dynamic deformation density plots (± 0.1 eÅ^{-3}) and the colour scheme has been scaled to ± 1.0 eÅ^{-3}. The range of electron density, however, is larger than the dynamic densities (-0.48 to +2.42 eÅ^{-3}). This is to be expected as atomic displacements lead to smearing of the density which results in the smaller values for electron density maxima in the dynamic deformation density plot.[138,139] The shape of the contours of electron density within the bonds differ between the two plots, with the CN triple bonds showing two peaks of electron density, which has also been reported in published comparisons of dynamic and static electron densities.[138]

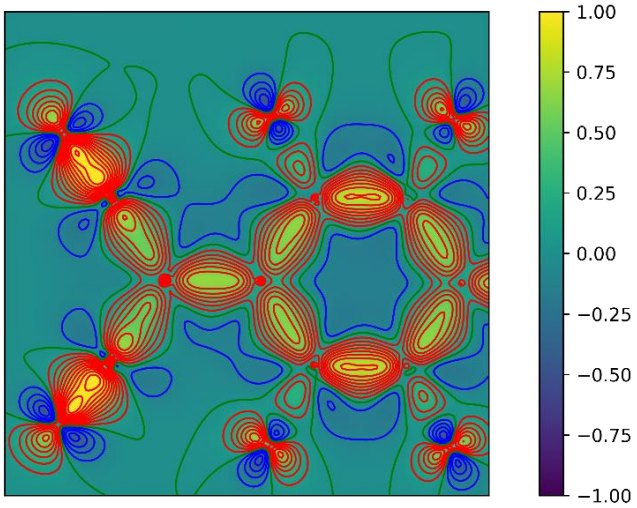


Figure 3.11 Static deformation electron density for reference model. Contours are recorded at levels of ± 0.10 eÅ^{-3}, where red indicated positive contours and blue indicates negative. The zero contour is pictured in green. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

Atom 1	Atom 2	Bond length (Å)
F(1)	C(1)	1.3271(3)
F(2)	C(2)	1.3261(3)
N(1)	C(5)	1.1577(3)
N(2)	C(6)	1.1593(4)
C(1)	C(3)	1.4400(3)
C(2)	C(3)	1.4398(3)
C(1)	C(2*)	1.3515(3)
C(3)	C(4)	1.3768(3)
C(4)	C(5)	1.4287(3)
C(4)	C(6)	1.4296(3)

*Table 3.5 Bond lengths as calculated by XDGEOM for reference processing. * indicates the bond between a symmetry related atom not within the asymmetric unit.*

The bond lengths calculated from the multipole model (Table 3.5) have smaller standard uncertainties than those calculated for the IAM (Table 3.2) indicating they were calculated to a higher degree of precision. The same distribution of bond lengths is observed for the multipole model.

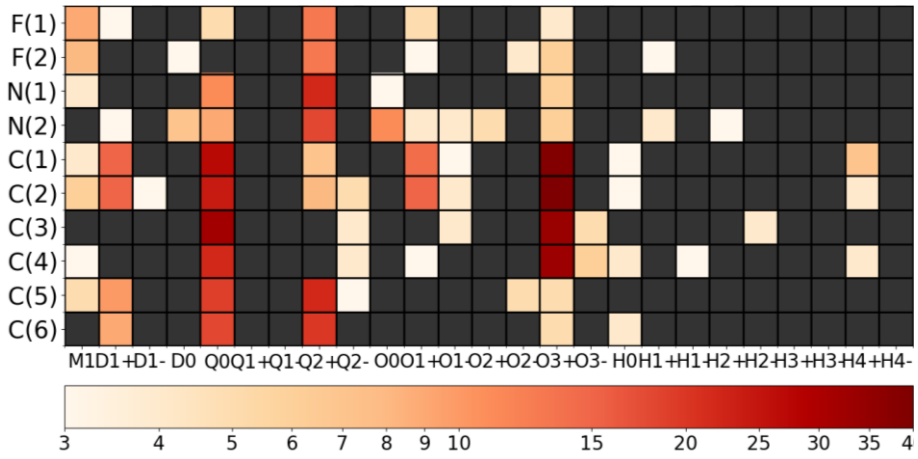


Figure 3.12 The significance of multipole parameters for the reference processing.

The significance of the multipole parameters is presented visually in Figure 3.12. Each grid square shows the significance of a parameter, calculated using X/σ_X, where X is the multipolar parameter and σ_X is the standard uncertainty of the parameter. This plot shows how many of the multipolar parameters have statistically significant values i.e. are significantly different from their unrefined values (e.g. 0 or the expected valence electron population for the monopole population M1). Those parameters which are not statistically significant ($X/\sigma_X < 3$) are shown in grey, with the rest of the grid squares coloured according to a logarithmic colour scheme. Amongst other features, this plot reveals that the population of some hexadecapoles is statistically significant. This is unexpected, as hexadecapoles have the same shape as f molecular orbitals and may suggest some non-classical bonding character present in the model.

Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.135(15)	-23.381(85)	0.1
F(2)	C(2)	2.096(15)	-21.754(86)	0.08
N(1)	C(5)	3.397(25)	-28.932(199)	0.06
N(2)	C(6)	3.373(24)	-29.683(201)	0.06
C(1)	C(3)	2.011(8)	-16.645(30)	0.18
C(2)	C(3)	2.009(9)	-17.119(30)	0.16
C(1)	C(2*)	2.391(26)	-23.629(67)	0.39
C(3)	C(4)	2.209(10)	-20.075(35)	0.27
C(4)	C(5)	1.955(9)	-14.829(31)	0.11
C(4)	C(6)	1.913(9)	-13.004(32)	0.08

Table 3.6 Properties at (3,-1) Bond critical points (BCPs) for reference processing.

Table 3.6 shows the density at BCPs for intramolecular bonds (between 1.1 – 1.6 Å). The amount of electron density at the BCP, ρ, generally decreases with decreasing bond strength. The value is largest for CN bonds and smallest for C – C single bonds; this trend is also mirrored in the Laplacian $\nabla^2\rho$ values. The values for ellipticity, ε, suggest conjugation within the molecule, as the values for the single bonds are elevated from 0. The largest values of ε are for the carbon double bonds: C(4)-C(3) and C(1) – C(2*), with the ε for C(1) – C(2*) close to that expected for a benzene double bond.[140] The ellipticity for C(3) – C(4) bond is smaller than expected for a double bond, possibly due to the conjugation within the molecule.

Atom	P_{val}	κ	κ’	Net atomic charge	Stockholder charges
F(1)	7.101(11)	0.999	0.953	-0.100(11)	-0.07376
F(2)	7.097(11)	0.999	0.953	-0.096(11)	-0.06679
N(1)	5.108(25)	0.996(2)	1.009	-0.108(25)	-0.1221
N(2)	5.058(26)	0.996(2)	1.009	-0.057(26)	-0.08904
C(1)	3.908(19)	1.016(2)	0.985	+0.092(19)	0.078555
C(2)	3.883(19)	1.016(2)	0.985	+0.117(19)	0.084676
C(3)	3.982(21)	1.010(2)	0.979	+0.017(21)	0.000958
C(4)	4.084(22)	1.010(2)	0.979	-0.083(22)	0.023932
C(5)	3.856(28)	1.039(2)	0.990	+0.143(28)	0.109528
C(6)	3.924(29)	1.039(2)	0.990	+0.075(29)	0.053766

Table 3.7 Atomic charges for reference processing as calculated by XDGEOM, errors for Kappa for F(1) and F(2) are too small to report by XDPROP. There are no errors reported for κ’ parameters as these parameters are not refined in final iteration of least squares.

The net atomic charges can be calculated by *XDGEOM* (Table 3.7) from the refined valence populations of the atoms. The charges on the atom approximately follow the order of electrostatics apart from N(1), which has the most negative charge, to C(4), where there is a strong negative charge. Some atoms which would be expected to have similar values due to their similar chemical environment appear to have different charges e.g. N1/N2 vs C5/C6, but these differences are not statistically significant. The most positive charges are recorded

for the carbon atoms connected to electron withdrawing atoms (C1, 2, 5 and 6). Net charges using the Stockholder routine in *XDPROP* were also calculated.[141] These have a general good agreement with net atomic charges, which is expected as they are calculated using a similar partitioning of electron density. However, the value for C(4), which is not within the error reported for the net charge, is calculated as positive for the Stockholder method.

The distribution of charge within the asymmetric unit, as calculated from the refined valence populations, is illustrated in Figure 3.13. This shows alternating positive and negatively charged atoms within the malononitrile group, which could be explained classically by electrostatics. C(3) is positively charged, but this value is small and not statistically significant. This could be due to the counteracting effect of the positive C(1) and C(2) vs the negative C(4). Within the ring of the molecule, all carbon atoms are positively charged, which would make the quadrupolar moment of ring π system in the z direction (above and below the ring) partially positively charged (as opposed to negative charges as found in TCNQ or benzene).[142,143] This has impact on the crystallisation of F_4TCNQ in toluene, as described in Chapter 4.

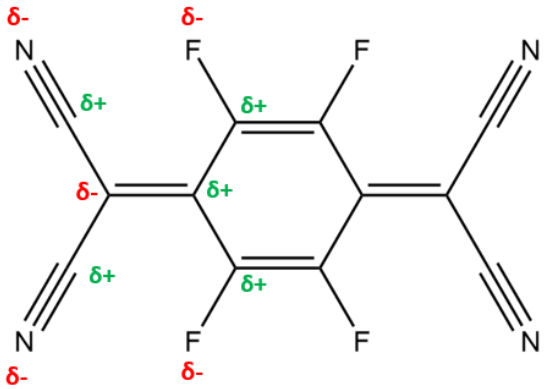


Figure 3.13 Distribution of net atomic charge in asymmetric unit from reference processing.

The assessment of the multipole model and the properties derived from it shows that the result of the refinement is chemically reasonable. While there is a possible feature in the residual electron density around N(2), this refinement will be used as the reference refinement. Due to the choice of mainly default processing options, this model could be improved upon; the reference refinement should not to be mistaken for the optimised refinement and is used here purely for comparison.

3.3 Considerations for comparisons of multipole models

If the way in which data are processed affects the completed multipole model, then comparisons between models in which different processing options have been used would

be restricted to qualitative analysis only. For this study, a number of different data processing options were varied and the final models produced were compared to the reference model (or another model where appropriate) in order to assess whether multipolar parameters and properties calculated from the models were statistically different from each other. A number of different refinement properties are investigated as listed below:

- multipole parameters,
- density at bond critical points,
- net charges on molecules,
- where appropriate, differences in *.hkl* files.

These properties were chosen as they have quantitative values and are commonly reported in charge density refinements. Many of these values also have associated standard error values which allows comparison of parameters of different models to take place with respect to the uncertainty of the measurement. Values were deemed statistically different from each other if the range of ±3σ of a value does not intersect with the range of ±3σ of another value or $\Delta > 3(\sigma_1 + \sigma_2)$ i.e. the difference between the two values was greater than 3 times the sum total of both of the standard uncertainty values.

Each model will also be assessed using normal probability plots, residual density plots and residual density analysis to ensure that the model is chemically reasonable. As far as possible the parameter list, i.e. the order in which the parameters were included in the multipolar refinement, was kept the same as the reference parameter list. However, this was not always achievable as the refinement can be sensitive to the order that parameters are included, as well as which parameters are refined simultaneously. Some combinations of processing choices would not converge using the same refinement strategy as the reference model and therefore it was necessary to modify the strategy. This could be a source of differences between models, although this effect is expected to be minimal.

Dipole moments will not be considered at this stage as no standard uncertainties reported for these parameters (as they are not included in the current version of *XD2016*). This makes them a potentially unreliable indicator, although they have been used for comparison in other studies. Figures and values will be reported as necessary; further results and graphs can be found in the Appendix.

3.4 Variation of multipole refinement options

There are many different parameters which can be applied in the process of multipolar refinement. Additionally, data can be excluded from the refinement by applying cut-offs. The order and choice of multipole parameters that are refined as well as the values of cut-offs can affect whether a refinement will converge and thus the outcome of the refinement. Investigations into the effect of different multipolar processing routines (or refinement strategies) on the final multipole models produced are described below. Multipole refinements were undertaken using same starting input files and *.hkl* files that were created during the reference processing.

3.4.1 Refinement of κ' parameter

The *XD* manual describes κ', the expansion or contraction of multipole parameters, as a very sensitive parameter and advises the comparison of final models with and without the κ' included.[43] The κ' parameter is usually included and refined alone in the modelling process. All other parameters are refined separately after convergence of the κ' parameters, as correlation with other parameters often leads to the least squares minimisation being unstable, preventing the convergence of the model. However, this process (called block refinement) can introduce bias into the refinement as not all parameters are being refined at once.[37] In the reference model, the κ' parameter is refined independently after convergence of the refinement with *xyz*, U_{ij}, multipoles to hexadecapole level and κ parameter. The differences in the final models were compared in multipole refinements where κ' is not refined (the value for κ' remains 1), block refined once (as in the reference refinement), and block refined twice. Upon further refinement of κ' using the above method, the model parameters did not converge when all parameters except κ' were refined.

The refinement statistics for the three models are reported in Table 3.8, including the values from residual density analysis[37] ($d^f(0)$, e_{net}, e_{gross}) for a 7x7x3.5 Å^3 grid surrounding the asymmetric unit. These parameters give a numerical value to the 'featureless-ness' of the residual electron density and comparing the parameters for the same grid in each model allows differences in the residual density surrounding the asymmetric unit to be assessed. The decrease in e_{net} and e_{gross} with increasing iterations of κ' refinement shows a reduction in the unmodelled density and suggests an improved model of the electron density. The increase in $d^f(0)$ can also be attributed to a decrease in the features of the residual electron density.[37] The e_{hole} and e_{peak} from a 3D residual electron density map are reported for the

same grid to assess if the range of residual electron density is changing from one model to the next; the range is approximately the same for all the three models.

	No κ'	Reference	κ' twice
R_2 (%)	1.59	1.55	1.53
R_1 (%)	1.49	1.48	1.47
wR_2 (%)	1.74	1.71	1.70
GooF	1.9986	1.9668	1.9446
$d^f(0)$	2.4844	2.4853	2.4855
e_{net} (e)	-0.0322	-0.0320	-0.0267
e_{gross} (e)	1.6678	1.6545	1.6476
e_{hole} ($eÅ^{-3}$)	-0.174	-0.174	-0.176
e_{peak} ($eÅ^{-3}$)	0.207	0.204	0.204

Table 3.8 Refinement statistics for κ' investigations.

There are very slight changes in the refinement statistics with increasing refinement of κ', which could imply the improvement of the model with a reduction in the R factors and GooF values. Normal probability plots produced for the models exhibit the same form as the reference refinement (Figure 3.9) with a non-normal distribution of residuals.

A comparison of the values of the multipolar parameters shows no statistical difference between the multipole parameters of the three models, which can therefore be classed as transferable. Properties are now calculated from each model to assess any differences in the values, even though the multipolar parameters are ostensibly the same.

	no κ'				κ' twice			
Atom	P_{val}	κ	κ'	Net atomic charge	P_{val}	κ	κ'	Net atomic charge
F(1)	7.103(11)	0.999	1	-0.103(11)	7.100(11)	0.998	0.916	-0.099(11)
F(2)	7.098(11)	0.999	1	-0.098(11)	7.096(11)	0.998	0.916	-0.096(11)
N(1)	5.125(25)	0.996(2)	1	-0.124(25)	5.093(25)	0.997(2)	1.018	-0.092(25)
N(2)	5.072(26)	0.996(2)	1	-0.072(26)	5.044(25)	0.997(2)	1.018	-0.044(25)
C(1)	3.910(19)	1.018(2)	1	+0.090(19)	3.905(20)	1.014(2)	0.973	+0.095(20)
C(2)	3.885(18)	1.018(2)	1	+0.114(18)	3.880(19)	1.014(2)	0.973	+0.120(19)
C(3)	3.979(21)	1.012(2)	1	+0.021(21)	3.987(22)	1.008(2)	0.962	+0.013(22)
C(4)	4.077(21)	1.012(2)	1	-0.077(21)	4.088(22)	1.008(2)	0.962	-0.087(22)
C(5)	3.841(29)	1.041(3)	1	+0.158(29)	3.870(28)	1.037(2)	0.981	+0.130(28)
C(6)	3.909(29)	1.041(3)	1	+0.091(29)	3.938(29)	1.037(2)	0.981	+0.061(29)

Table 3.9 Charges for molecule calculated for no κ' refinement and κ' twice.

There are no statistical differences between the calculated net atomic charge of atoms within each model (Table 3.7 and Table 3.9) and the differences between the values are extremely small. However, increasing the number of times κ' is refined generally results in a κ' value that is further from 1, and a value of κ that is closer to 1. The net atomic charges,

calculated from the electron density on each molecule (P_{val}), are also generally reduced (closer to neutral, 0) when κ' is refined. This results in the net charge for C(6) no longer being statistically significant in the model where κ' is refined twice. The order of the relative charge of the atoms also changes, which could impact any electrostatic properties calculated from the model, e.g. dipole moments. The difference between the net atomic charge of N(1) and N(2) and C(1) and C(2), however, is maintained across the models.

		no κ'			κ' twice		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.145(16)	-23.971(91)	0.1	2.126(15)	-22.899(81)	0.1
F(2)	C(2)	2.105(16)	-22.264(93)	0.08	2.089(15)	-21.328(82)	0.08
N(1)	C(5)	3.390(27)	-28.197(216)	0.06	3.402(23)	-29.573(186)	0.06
N(2)	C(6)	3.369(27)	-29.206(218)	0.07	3.376(22)	-30.087(188)	0.06
C(1)	C(3)	2.011(9)	-16.448(33)	0.19	2.010(8)	-16.774(27)	0.18
C(3)	C(2)	2.010(9)	-16.961(33)	0.16	2.008(8)	-17.217(28)	0.16
C(1)	C(2*)	2.390(27)	-23.288(70)	0.4	2.392(26)	-23.863(64)	0.39
C(3)	C(4)	2.209(10)	-19.786(39)	0.28	2.209(9)	-20.271(33)	0.26
C(4)	C(5)	1.955(10)	-14.612(34)	0.11	1.956(9)	-14.993(29)	0.11
C(4)	C(6)	1.912(10)	-12.783(35)	0.08	1.913(9)	-13.171(30)	0.08

Table 3.10 Properties at BCPs for no κ' refinement and 2 κ' refinements.

The properties calculated at BCPs can also be compared (Table 3.6 and Table 3.10). There is no statistical difference in the density at the BCP when varying the κ' refinement with respect to those of the reference model. However, there is a trend in the values at the BCPs that when the number of times κ' is refined is increased the value of ρ(r) also increases for all bonds except F-C. This trend is repeated and exaggerated for the Laplacian values, resulting in a small number of values having statistically significant differences in Laplacian values (C(3) – C(4) and C(4) – C(5) for both models). The ellipticity calculated at the BCP is essentially the same across all models.

Increasing the number of times κ' is refined in this case improves refinement statistics, although only marginally when the I/σ_I cut-off is 3. For other processing options, refining κ' more than once prevents the model from converging. It was seen that a refinement with no I/σ_I cut-off would only converge when this parameter was refined twice, thereby demonstrating the sensitivity of the multipole model to the refinement of this parameter.

3.4.2 Variation of the I/σ_I cut-off

The default option in *XD* is to exclude all reflections where the I/σ_I value is less than 3; common values in literature for this cut-off when applied in charge density refinements are

3 and 2.[88] This can be justified as the intensity values for these reflections are not statistically significant above zero, although the removal of less significant values has been shown to introduce bias into the distribution of the residuals of the model which prevents them from being classed as truly random values (as the distribution of residuals is now dependent on factors such as the resolution of the reflection or intensity value).[42] Convergence of some refinement strategies was dependant on the value of the cut-off and hence this is another important refinement option to investigate.

The initial *XD* inputs (*.hkl*, *.mas*, *.inp*) from the reference processing routine were taken as a start point for the refinement and the same refinement strategy was used up to the inclusion of κ'. As stated in the previous section, the refinement would only converge when the cut-off was removed completely if κ' was refined twice before the I/σ_I cut-off was removed, therefore the additional refinement of κ' was included in the strategy. The refinement would also not converge if there was no initial I/σ_I cut-off included at the start. After convergence of the model with an I/σ_I cut-off of < 3, the cut-off was removed in steps of 1 until the desired cut-off was reached. The *xyz*, U_{ij}, multipolar and κ parameters were then refined together to produce the final model. The multipolar refinement strategy utilised is described in Section 10.1.3.2. Due to the refinement strategy used, the completed models are therefore compared to that of the model where κ' was block refined twice from the previous section instead of the reference model, in order to provide a more accurate picture of the effect of removing the I/σ_I cut-off on the final model.

	κ' twice	I/σ_I <2	I/σ_I <0
# Reflections	4029	4201	4670
R_2 (%)	1.53	1.54	1.57
R_F (%)	1.47	1.58	1.99
wR_2 (%)	1.70	1.71	1.75
GooF	1.9446	1.9228	1.8514
$d^f(0)$	2.4855	2.4909	2.5059
e_{net} (e)	-0.0267	-0.0280	-0.0303
e_{gross} (e)	1.6476	1.7559	2.1645
e_{hole} ($eÅ^{-3}$)	-0.176	-0.181	-0.232
e_{peak} ($eÅ^{-3}$)	0.204	0.221	0.234

Table 3.11 Refinement statistics for I/σ_I variation.

When comparing the refinement statistics of the completed models (Table 3.11), the immediate effect of reducing the I/σ_I cut-off is an increase in the number of reflections included in the refinement. There is also an increase in the noise observed in the refinement, shown in the increase in range of residual electron density, which may be expected as these

extra reflections have low I/σ_I values. This may also be the cause of the increase in the R factors. The GooF, however, improves with the inclusion of these less significant reflections, although the value is still far from ideal.

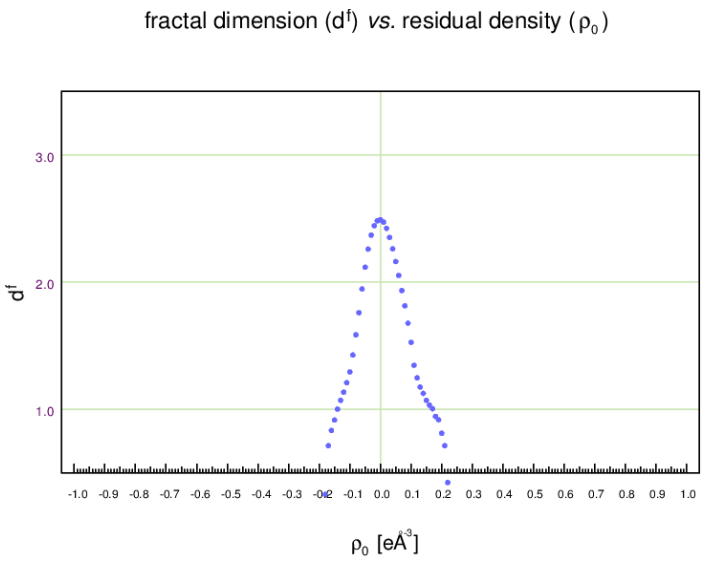


Figure 3.14 Fractal dimensionality plot for model with $I/\sigma_I < 2$ cut-off.

Similarly to the previous section, there is no statistical difference between values for multipolar parameters between the three refinements and therefore multipolar parameters are again deemed transferable between models. All models pass Hirshfeld rigid bond test and produce a non-normal distribution of structure factor residuals with a similar shape to that of the reference processing (Figure 10.6). Fractal dimensionality plots show the formation of a slight 'shoulder' in the positive residual density, which can occur when there are errors in κ refinement (visible in Figure 3.14 and Figure 10.8).[37]

	κ' twice			$I/\sigma_I < 2$			no I/σ_I cut-off		
	P_{val}	κ	Net atomic charge	P_{val}	κ	Net atomic charge	P_{val}	κ	Net atomic charge
F(1)	7.100(11)	0.998	-0.099(11)	7.099(11)	0.999	-0.098(11)	7.098(11)	0.998	-0.097(11)
F(2)	7.096(11)	0.998	-0.096(11)	7.095(11)	0.999	-0.094(11)	7.094(10)	0.998	-0.093(10)
N(1)	5.093(25)	0.997(2)	-0.092(25)	5.091(24)	0.997(2)	-0.090(24)	5.084(23)	0.997(1)	-0.083(23)
N(2)	5.044(25)	0.997(2)	-0.044(25)	5.044(25)	0.997(2)	-0.044(25)	5.040(24)	0.997(1)	-0.040(24)
C(1)	3.905(20)	1.014(2)	+0.095(20)	3.905(19)	1.014(2)	+0.094(19)	3.905(19)	1.014(2)	+0.095(19)
C(2)	3.880(19)	1.014(2)	+0.120(19)	3.881(19)	1.014(2)	+0.119(19)	3.882(18)	1.014(2)	+0.118(18)
C(3)	3.987(22)	1.008(2)	+0.013(22)	3.987(22)	1.008(2)	+0.012(22)	3.986(21)	1.008(2)	+0.014(21)
C(4)	4.088(22)	1.008(2)	-0.087(22)	4.088(22)	1.008(2)	-0.088(22)	4.087(21)	1.008(2)	-0.087(21)
C(5)	3.870(28)	1.037(2)	+0.130(28)	3.873(28)	1.037(2)	+0.127(28)	3.881(27)	1.037(2)	+0.118(27)
C(6)	3.938(29)	1.037(2)	+0.061(29)	3.939(28)	1.037(2)	+0.061(28)	3.944(27)	1.037(2)	+0.055(27)

Table 3.12 Net atomic charges for I/σ_I cut-off variation.

There is no statistical difference in the net atomic charge between the three refinements (Table 3.12), although there are variations in the values of the parameters. Again, there is a general trend identified across these parameters, in this case the value of the net atomic

charge decreases with the inclusion of lower significance reflections. κ' values are not included as they are the same for all models – as the κ' value is only refined when the $I/\sigma_I < 3$ cut-off is included since convergence was not achieved if it was refined when the I/σ_I cut-off was removed completely.

		$I/\sigma_I < 2$ cut-off			no I/σ_I cut-off		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.127(14)	-22.886(79)	0.1	2.125(14)	-22.901(76)	0.1
F(2)	C(2)	2.090(14)	-21.306(80)	0.08	2.091(14)	-21.326(77)	0.08
N(1)	C(5)	3.404(22)	-29.737(182)	0.06	3.410(21)	-30.690(173)	0.06
N(2)	C(6)	3.376(22)	-30.117(184)	0.06	3.380(21)	-30.633(176)	0.06
C(1)	C(3)	2.010(8)	-16.783(27)	0.18	2.010(8)	-16.796(26)	0.18
C(2)	C(3)	2.009(8)	-17.232(28)	0.16	2.008(8)	-17.226(27)	0.16
C(1)	C(2*)	2.392(25)	-23.875(63)	0.39	2.392(24)	-23.901(60)	0.39
C(3)	C(4)	2.209(9)	-20.270(32)	0.26	2.209(8)	-20.295(31)	0.26
C(4)	C(5)	1.954(9)	-14.967(29)	0.1	1.955(8)	-14.998(28)	0.1
C(4)	C(6)	1.913(9)	-13.174(29)	0.08	1.913(8)	-13.203(28)	0.08

Table 3.13 Properties at BCPs for I/σ_I cut-off variation.

There are no statistical differences in the value of ρ at the BCPs across the three models (Table 3.13) and similar to the previous section, the ellipticities calculated at the BCPs are the same for all models. There is also no statistical difference in the Laplacian values at BCPs, except for that of N(1) – C(5). The difference in the multipole parameters divided by the sum of the standard uncertainties of both values is 3.11 and therefore the difference in the values is only slightly above the criteria for statistical difference.

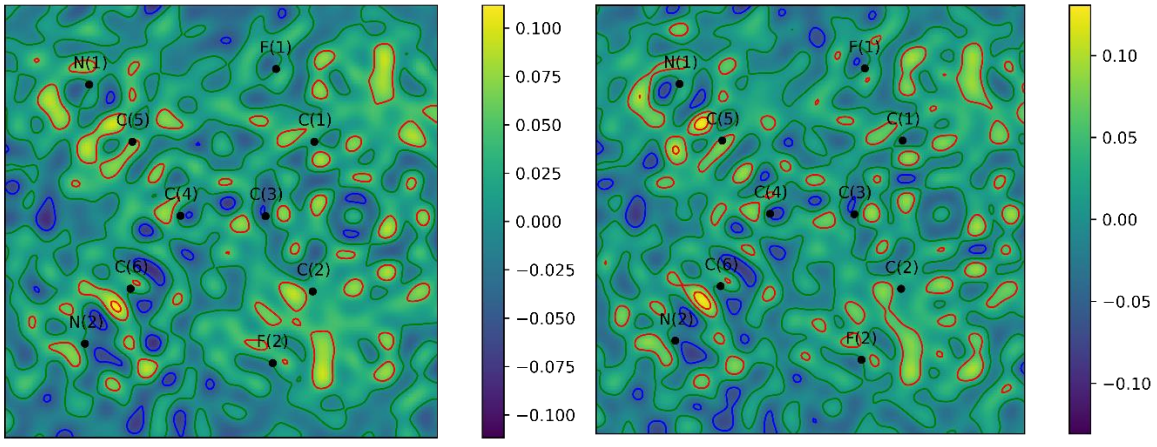


Figure 3.15 Residual electron density plots for $I/\sigma_I < 2$ and $I/\sigma_I < 0$. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green.

In this case, there is very little difference between the models produced after the I/σ_I cut-off is removed. This may be because the removal of the cut-off happens after the refinement has converged and κ' has been refined, potentially biasing the model towards that which

was created when the $I/\sigma_I < 3$ cut-off was employed. When the less significant reflections are removed, better values for the residual electron density and refinement statistics are reported as less of the experimental noise is included in the refinement. This can be seen when comparing the 2D residual density plots in the plane of the molecule (Figure 3.15) where removing the cut-off increases the range of the residual electron density in the plane of the molecule and increases the 'noise' in the background, as evidenced by the increase in the number and height of contours. This 'noise' may however be important information in the refinement which has not been modelled, as the largest peaks and troughs of the residual electron density lie around the CN bonds.

3.4.3 Chemical constraints

Chemical constraints can be applied to the refinement in order to constrain the multipolar parameters of atoms in similar chemical environments to the same values. This reduces the number of parameters to be refined and thereby increases the data to parameter ratio. Krause *et al.*[62] recommend the use of chemical constraints to prevent the overfitting of data.

The same starting point as the reference refinement was again used in this case. Chemical constraints were applied to N(2), C(2), F(2) and C(6) to fix the multipole parameters to those of N(1), C(1), F(1) and C(5) respectively. Convergence was not achieved using the reference refinement strategy and therefore the strategy had to be modified. After a number of refinement strategies were tested, converge of the multipolar parameters was only achieved when an $I/\sigma_I < 2$ cut off was applied throughout the refinement, instead of $I/\sigma_I < 3$ as used in the reference processing. The strategy applied to produce a final converged multipole model is given in Section 10.1.3.3; it uses the same parameter list as κ' refined twice (Section 3.4.1), with the exception of the $I/\sigma_I < 2$ cut-off. A corresponding unconstrained refinement, using the same parameter list and I/σ_I cut-off, was also undertaken for comparison against the constrained refinement to minimise the sources of any discrepancy in parameters.

Refinement statistics for the final multipole models are listed in Table 3.14. In most cases, the statistics for the unconstrained refinement are better than those for the constrained refinement; R factors, GooF and the range of residual electron density are all smaller for the unconstrained refinement and imply a better fit of the model to the data when the constraints are not applied. The e_{gross} value, which describes the total residual electron density within the analysed area, is very similar for both models indicating a similar model

error. The data to parameter ratio is also given and shows the expected increase in the ratio when the multipolar parameters are constrained.

	Unconstrained	Constrained
R_2 (%)	1.54	1.61
R_1 (%)	1.58	1.67
wR_2 (%)	1.72	1.87
GooF	1.9229	2.0652
$d^f(0)$	2.4909	2.4933
e_{net} (e)	-0.0280	-0.0451
e_{gross} (e)	1.7560	1.7494
e_{hole} (eÅ^{-3})	-0.181	-0.225
e_{peak} (eÅ^{-3})	0.221	0.274
N_{ref}/N_v	12.1768	17.1469

Table 3.14 Refinement statistics for unconstrained and constrained multipole models.

When assessing the quality of the models through other measures, normal probability plots for both models show a non-normal distribution of residuals (Figure 10.9). The fractal dimensionality plots show a less parabolic distribution of values for the chemically constrained refinement, signifying that the residual electron density within the asymmetric unit contains more features (Figure 3.16).

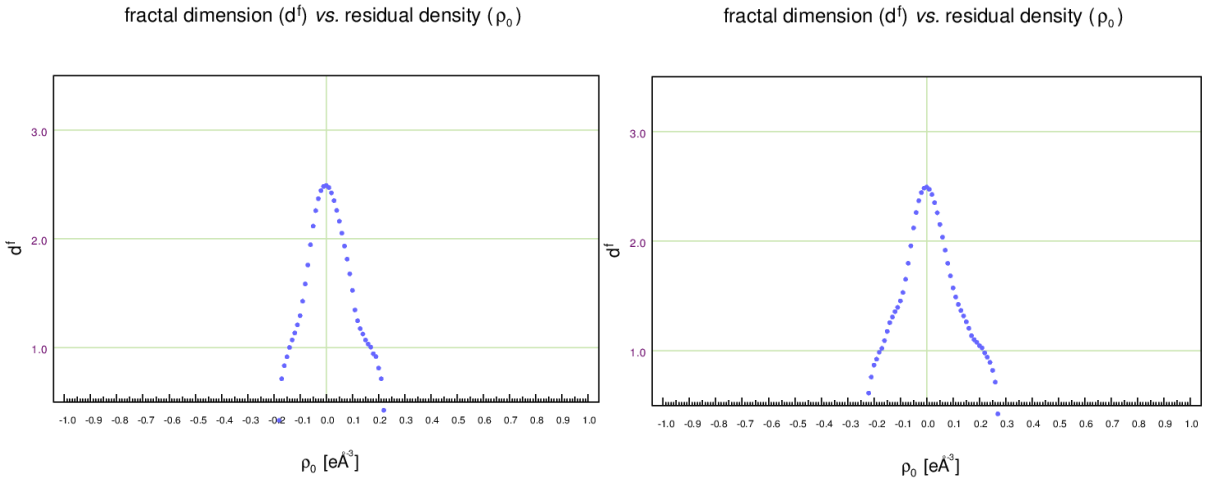


Figure 3.16 Fractal dimensionality plot for unconstrained (left) and constrained (right) multipole model.

Several multipole parameters are statistically different between the final unconstrained and constrained multipole model. The parameters in question are illustrated in Figure 3.17, where each grid square is related to one refined multipole parameter. In a similar fashion to Figure 3.12, the significant difference in the values of the multipole parameters in the two models, S_d, is calculated using Equation 3.1, where M is the value of the multipolar parameter. Those with a difference of greater than 2 are shown in blue, with the colour of the grid square relating to the total number of standard deviations that separate the two

values. Parameters where $S_d < 2$ are set to 0 in order to better illustrate the statistically significant (or nearly significant) parameters.

$$S_d = \frac{|M_1 - M_2|}{\sigma_2 + \sigma_2} \qquad 3.1$$

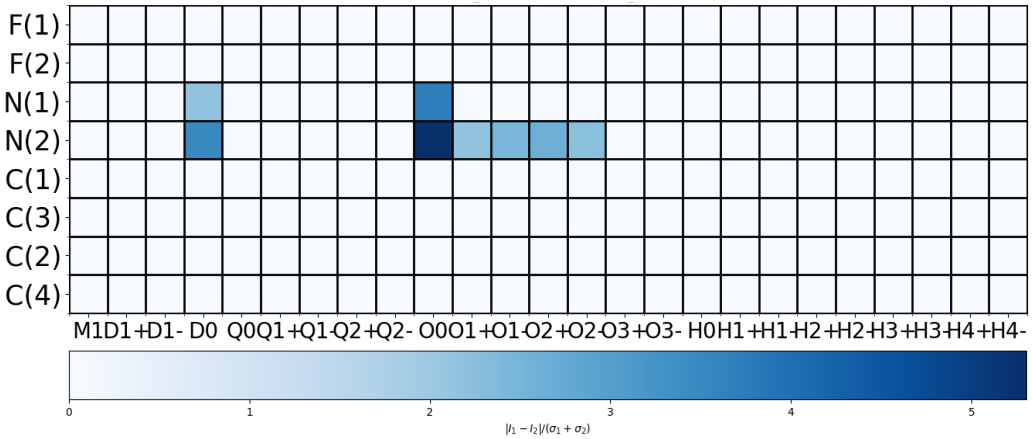


Figure 3.17 Multipolar parameters that are statistically significant between chemically constrained and reference models.

Parameters N(1) O(0), N(2) O(0) and N(2) D(0) are all statistically different between the two models, with an S_d of > 3. This may imply that applying chemical constraints to the N atoms is not reasonable, even though they appear to be in chemically equivalent environments. Possible explanations for this could be a difference in intermolecular interactions occurring from the two atoms or due to the accumulation of residual density around N(2) (see Section 3.2.5) leading to a difference in environments between the two atoms for these data.

	Unconstrained				Constrained			
Atom	P_{val}	κ	κ'	Net atomic charge	P_{val}	κ	κ'	Net atomic charge
F(1)	7.098(11)	0.999	0.917	-0.098(11)	7.088(11)	0.998(1)	0.908	-0.087(11)
F(2)	7.095(11)	0.999	0.917	-0.094(11)	7.088	0.998(1)	0.908	-0.087
N(1)	5.090(24)	0.997(2)	1.018	-0.090(24)	5.072(22)	0.995(2)	1.003	-0.072(22)
N(2)	5.044(25)	0.997(2)	1.018	-0.043(25)	5.072	0.995(2)	1.003	-0.072
C(1)	3.905(19)	1.014(2)	0.973	+0.095(19)	3.900(15)	1.013(2)	0.97	+0.099(15)
C(2)	3.880(19)	1.014(2)	0.973	+0.119(19)	3.9	1.013(2)	0.97	+0.099
C(3)	3.987(22)	1.008(2)	0.962	+0.013(22)	3.964(23)	1.006(2)	0.961	+0.036(23)
C(4)	4.088(22)	1.008(2)	0.962	-0.087(22)	4.069(23)	1.006(2)	0.961	-0.069(23)
C(5)	3.873(28)	1.037(2)	0.981	+0.127(28)	3.923(26)	1.034(3)	0.983	+0.077(26)
C(6)	3.939(28)	1.037(2)	0.981	+0.061(28)	3.923	1.034(3)	0.983	+0.077

Table 3.15 Net atomic charges calculated for chemically constrained refinement. Standard uncertainties are not recorded for P_{val} and net atomic charges of chemically constrained atoms.

There are no statistical differences between the net atomic charges of any atom between the two models (Table 3.15). Furthermore, there are no statistical differences between the total density at BCPs between the two models (Table 3.16) and ellipticities at the BCPs remain approximately the same. Some values for the Laplacian are statistically different as

may be expected when comparing a relatively sensitive parameter; interestingly the values for $\nabla^2\rho$ for N(1) and N(2) are very similar in both models, as was observed with a model in which no I/σ_I cut-off was applied. The cause of this similarity is unknown.

		Unconstrained			Constrained		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.125(14)	-22.901(76)	0.1	2.111(11)	-22.150(61)	0.09
F(2)	C(2)	2.091(14)	-21.326(77)	0.08	2.116(1)	-22.378(0)	0.009
N(1)	C(5)	3.410(21)	-30.690(173)	0.06	3.401(19)	-31.067(149)	0.06
N(2)	C(6)	3.380(21)	-30.633(176)	0.06	3.399(0)	-31.072(0)	0.06
C(1)	C(3)	2.010(8)	-16.796(26)	0.18	2.017(7)	-17.355(25)	0.18
C(2)	C(3)	2.008(8)	-17.226(27)	0.16	2.003(12)	-17.300(35)	0.16
C(1)	C(2*)	2.392(24)	-23.901(61)	0.39	2.397(5)	-24.063(19)	0.39
C(3)	C(4)	2.209(8)	-20.295(31)	0.26	2.201(9)	-20.348(34)	0.25
C(4)	C(5)	1.955(8)	-14.998(28)	0.10	1.943(8)	-14.594(27)	0.09
C(4)	C(6)	1.913(8)	-13.203(28)	0.08	1.918(13)	-13.766(37)	0.09

Table 3.16 Properties at BCPs for unconstrained and chemically constrained refinement.

The inclusion of chemical constraints is said to prevent overfitting of the model – however contrary to expectation the N(1) and N(2) atoms appear to be chemically different, resulting in statistically different values for some multipolar parameters when compared to the equivalent unconstrained model. In order to assess if overfitting were a problem for this refinement, necessitating the application of constraints, a cross-validation refinement could be performed.

3.4.4 Weighting schemes

To obtain a more normal distribution of weights, some charge density refinements include a weighting scheme other than that of statistical weights. Values for weighting parameters for a *SHELXL* weighting scheme (Equation 1.24) have previously been calculated using trial and error to produce a more linear distribution of residuals in the normal probability plot or to return a wGooF of ~ 1.[38,95] Due to the frequency of this technique being utilised, the effect of applying a *SHELXL* weighting scheme on the multipole model was investigated.

CAPOW was used to calculate optimal values for the *a* and *b* weighting parameters for the completed reference refinement. Iterations of the calculation of values for *a* and *b* followed by least squares refinement of all parameters (except κ') with the new weighting scheme applied took place until convergence of the values for the *a* and *b* parameters occurred to 4 decimal places. For the initial iterations of least squares refinement with the inclusion of the *SHELXL* weighting scheme, κ parameters were also not refined. After ~ 3 iterations they were

then included in the refinement. Convergence of the values of *a* and *b* would not occur in refinements where κ parameters were included when the weighting scheme was initially applied. This problem occurred in most applications of the weighting scheme and could be due to correlations between parameters. The final values for the *a* and *b* in the converged *SHELXL* weighted model were *a* = 0.009 and *b* = 0.027.

	Reference	*SHELXL* weighted
R_2 (%)	1.55	1.56
R_F (%)	1.48	1.46
wR_2 (%)	1.71	2.75
wGooF	1.9668	0.9419
$d^f(0)$	2.4853	2.4863
e_{net} (e)	-0.0320	-0.366
e_{gross} (e)	1.6545	1.6006
e_{hole} (eÅ^{-3})	-0.174	-0.160
e_{peak} (eÅ^{-3})	0.204	0.198

Table 3.17 Refinement statistics for SHELXL weighted multipole model.

A comparison of the refinement statistics for the reference model, which has a statistical weighting scheme applied, and the *SHELXL* weighted model (Table 3.17) show that the R factors increase when a *SHELXL* weighting was applied, suggesting a worse fit of the calculated and observed structure factors. However, the wGooF, which is the same value as the GooF when the model only has a statistical weighting scheme applied, is much closer to the ideal value of 1 for the *SHELXL* weighted refinement. The range of residual electron density is also reduced for *a SHELXL* weighted refinement possibly due to the increase of the value used for the standard uncertainty of the parameters.

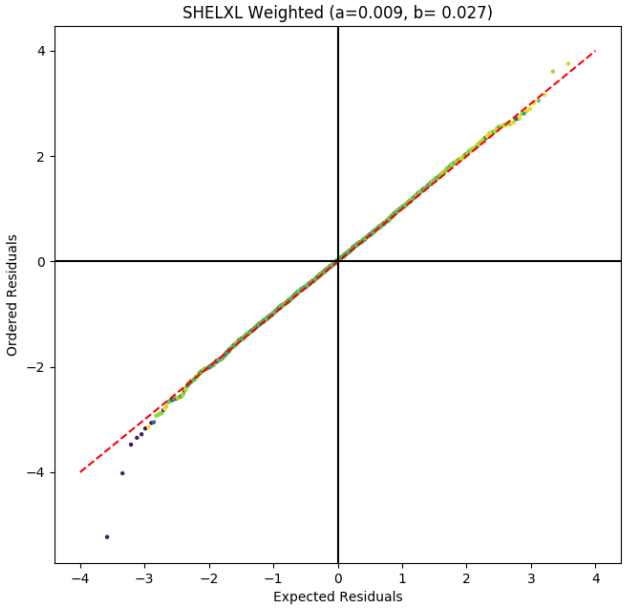


Figure 3.18 Normal probability plot calculated for reference refinement with an optimised SHELXL weighting scheme applied.

The normal probability plot for the *SHELXL* weighted model (Figure 3.18) shows an approximately normal distribution of values when compared to that of the reference refinement (Figure 3.9). There are a few outliers, identified as low angle reflections, which have a residual value that deviates from the normal distribution when the weighting scheme is applied, but these are smaller outliers than that of the statistically weighted model. The fractal dimensionality plot (Figure 10.12) has broadly the same shape as Figure 3.14 (when the I/σ_I cut-off is reduced) with a slight 'shoulder' in the positive residual density which could suggest some problem with the κ parameters.

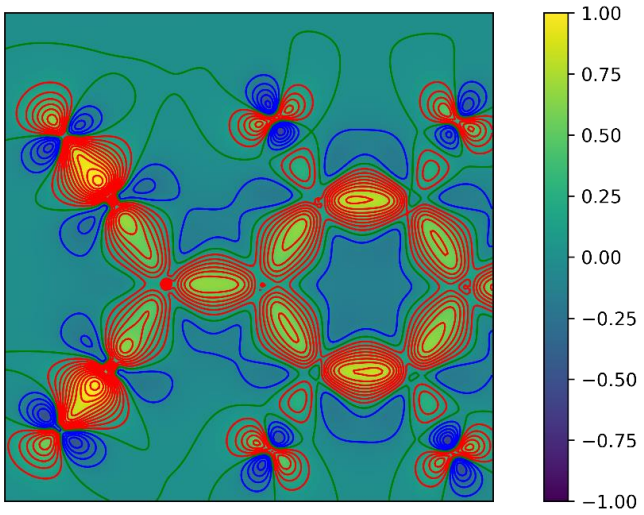


Figure 3.19 Static deformation density plot for reference data processing with SHELXL weighting applied. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

One difference between the two models can be seen in the static deformation density plot. The shape of the density between atoms differs when the weighting scheme is applied, exemplified by the change in the shape of the contours about the CN groups in Figure 3.19 (when compared to Figure 3.11). In order to investigate this further, the total density along the bond path of all bonds, using *XDPROP* (algorithm 6), was calculated. For the Laplacian or total density no appreciable difference was observed when the values for the two models were compared. However, an analysis of the ellipticity along the bond path shows that although the overall shape of the distributions is comparable – each bond has the same number of peaks and the positions of the peaks are approximately equivalent for both models – there were large deviations in the values of the ellipticity along C(1) – F(1) and C(6) – N(2) bonds (Figure 3.20 and further plots in section 10.1.4). Kaminski *et al.*[47] demonstrated an increasing deviation in the Laplacian and total density as the distance from the BCP increased in their analysis of oxalic acid, so the deviations could still be within an acceptable

range. The overall form of the ellipticity in all other bonds is much more similar to that of the reference refinement.

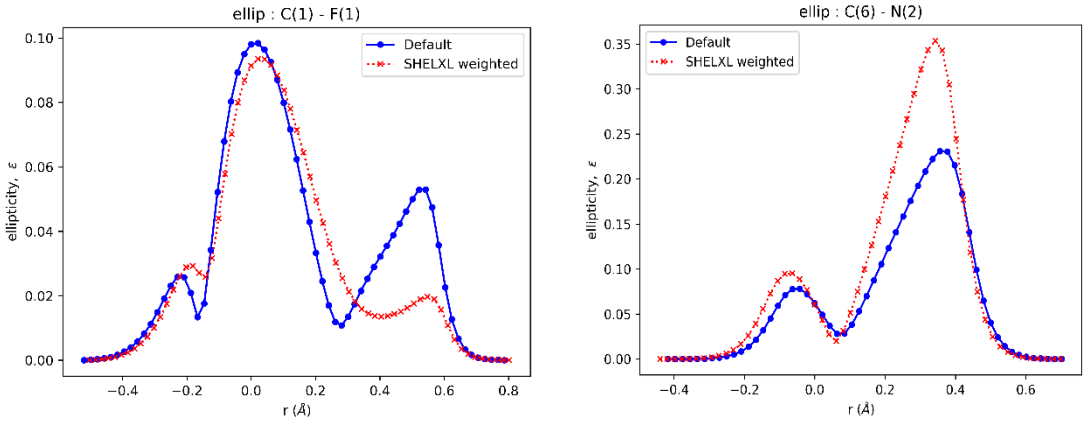


Figure 3.20 Ellipticity plotted along bond path for statistical and SHELXL weighted refinement.

The application of the *SHELXL* weighting scheme results in some multipolar parameters becoming less significant or no-longer statistically significant (when comparing Figure 3.12 and Figure 3.21), an effect which was also detailed in Chapter 2.

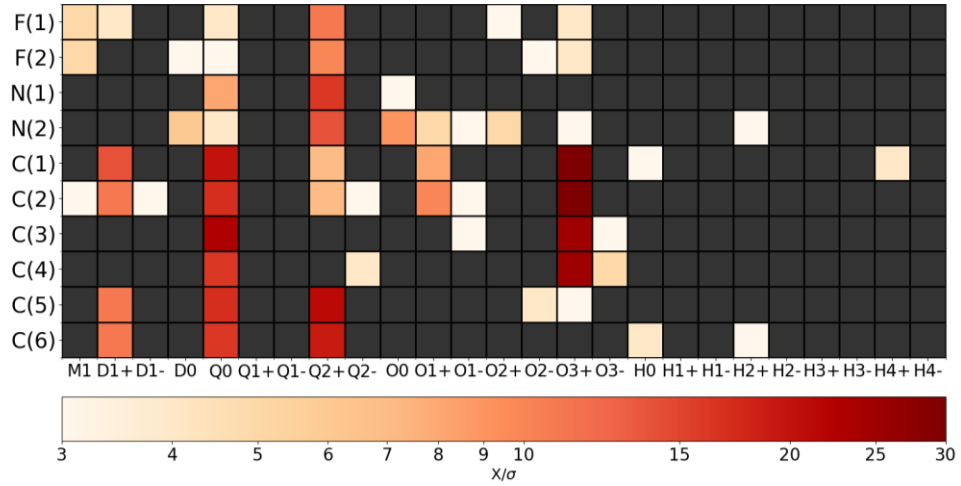


Figure 3.21 Statistical significance of multipolar parameters for SHELXL weighted model.

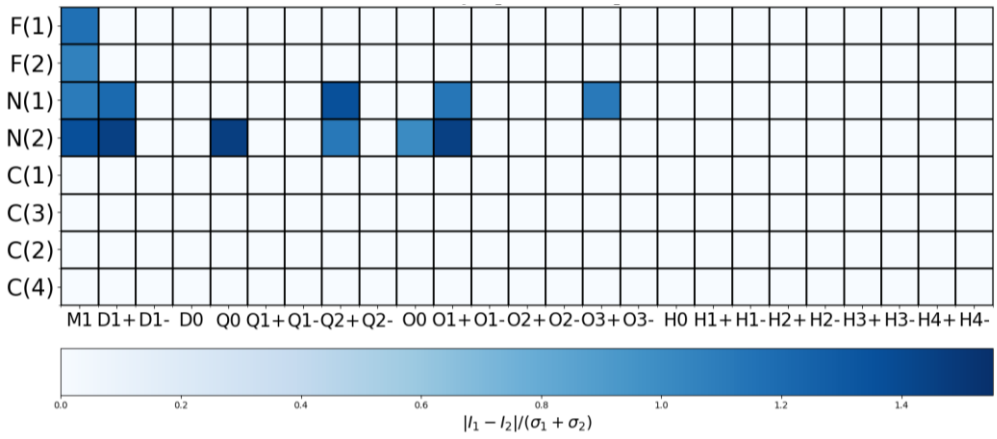


Figure 3.22 Differences between reference and SHELXL weighted refinement.

Although the application of a *SHELXL* weighing scheme produces no statistical difference in values of the multipolar parameters, there is a greater deviation in the absolute values. For all previous refinement comparisons, the significant difference, S_d (Equation 3.1), of all parameters was < 1. In this refinement, a number of parameters have an S_d of > 1, up to a maximum of ~ 1.5 as illustrated in Figure 3.22.

		Reference			*SHELXL* weighted		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.135(15)	-23.381(85)	0.1	2.088(19)	-23.384(107)	0.08
F(2)	C(2)	2.096(15)	-21.754(86)	0.08	2.076(20)	-21.971(109)	0.07
N(1)	C(5)	3.397(25)	-28.932(199)	0.06	3.435(28)	-36.598(215)	0.05
N(2)	C(6)	3.373(24)	-29.683(201)	0.06	3.409(27)	-37.311(200)	0.06
C(1)	C(3)	2.011(8)	-16.645(30)	0.18	2.005(13)	-16.834(45)	0.18
C(2)	C(3)	2.009(9)	-17.119(30)	0.16	1.994(13)	-17.171(46)	0.16
C(1)	C(2*)	2.391(26)	-23.629(67)	0.39	2.397(32)	-24.319(87)	0.39
C(3)	C(4)	2.209(10)	-20.075(35)	0.27	2.196(14)	-20.036(54)	0.28
C(4)	C(5)	1.955(9)	-14.829(31)	0.11	1.946(14)	-14.913(49)	0.11
C(4)	C(6)	1.913(9)	-13.004(32)	0.08	1.911(14)	-13.536(49)	0.08

Table 3.18 Properties at BCPs for SHELXL weighted refinement.

An analysis of values calculated at BCPs shows no statistical differences between ρ calculated at the BCP (Table 3.18). There is however a large statistical difference between the values for $\nabla^2\rho$ at the BCP of the CN bonds indicating a greater charge accumulation at these points for the *SHELXL* weighted refinement.

	Reference				*SHELXL* weighted			
Atom	P_{val}	κ	κ′	Net atomic charge	P_{val}	κ	κ′	Net atomic charge
F(1)	7.101(11)	0.999	0.953	**-0.100(11)**	7.072(14)	0.998(1)	0.953	**-0.071(14)**
F(2)	7.097(11)	0.999	0.953	**-0.096(11)**	7.071(14)	0.998(1)	0.953	**-0.070(14)**
N(1)	5.108(25)	0.996(2)	1.009	**-0.108(25)**	5.049(29)	0.993(2)	1.009	-0.048(29)
N(2)	5.058(26)	0.996(2)	1.009	-0.057(26)	4.983(29)	0.993(2)	1.009	+0.017(29)
C(1)	3.908(19)	1.016(2)	0.985	**+0.092(19)**	3.944(25)	1.008(3)	0.985	+0.056(25)
C(2)	3.883(19)	1.016(2)	0.985	**+0.117(19)**	3.904(24)	1.008(3)	0.985	**+0.095(24)**
C(3)	3.982(21)	1.010(2)	0.979	+0.017(21)	3.948(27)	1.008(3)	0.979	+0.052(27)
C(4)	4.084(22)	1.010(2)	0.979	**-0.083(22)**	4.059(27)	1.008(3)	0.979	-0.058(27)
C(5)	3.856(28)	1.039(2)	0.990	**+0.143(28)**	3.953(34)	1.030(3)	0.99	+0.047(34)
C(6)	3.924(29)	1.039(2)	0.990	**+0.075(29)**	4.018(34)	1.030(3)	0.99	-0.017(34)

Table 3.19 Net atomic charges for SHELXL weighted refinement. Statistically significant charges are emboldened.

While the net atomic charges calculated for the *SHELXL* weighted refinement are not statistically different (Table 3.19), the absolute value of the charges are reduced for many atoms. The standard uncertainty of the parameters is also increased when the *SHELXL* weighting scheme is applied, a trend which can also be identified in the properties at BCPs in

Table 3.17. These factors lead to a reduction in the number of charges which are statistically significant from zero (emboldened within Table 3.19). It is also noted that the absolute value for the charge of N(2) and C(6) is the reverse of the expected order, although these values are not statistically distinct.

In summary, applying a weighting scheme produces a more normal distribution of residual structure factors and improves the distribution of the residual electron density. The linear normal probability plot shows that the underestimation of the standard uncertainties can be somewhat corrected for through the application of an optimised *SHELXL* weighted scheme. The application of the *SHELXL* weighting scheme causes the standard uncertainties on all parameters to become larger, which causes a reduction in the number of statistically significant multipolar parameters and parameters calculated from the model.

It should be reiterated that it is not always appropriate to apply a weighting scheme and that, similar to the recommendations for the analysis of the κ' parameter, the refinement before and after application of a *SHELXL* weighting scheme should be assessed. As there are no statistical differences in multipolar values, charges or the density at bond critical points this would suggest that the application of the *SHELXL* scheme is reasonable in this case.

3.5 Varying of scaling options

There are a number of different corrections that can be applied during the scaling of the data which would be expected to affect the values of the intensities and their standard uncertainties. This in turn could affect the final multipole refinement. A number of the options and corrections that can be applied in the scaling program *SADABS* are investigated below.

3.5.1 Error models for standard uncertainty values

SADABS has a number of different error models which can be applied to the standard uncertainties for the purpose of correcting for instrument errors and scaling values between different experimental runs. The default error model (model 5 in *SADABS*) is to refine an individual scale value obtained for each experimental run, K, and a global value for g, which is the multiplier for an intensity dependant correction, to apply to the standard uncertainty to produce a corrected value (Equation 1.28). The default error model is used in the scaling of the reflection data for the reference refinement. It has been suggested that this model

may not to be suitable for use in a charge density refinement[79–81] and therefore four of the other available error models have been investigated:

- Error model 0: No error model (K = 1, g = 0).
- Error model 1: Only an intensity dependant correction. An overall value for g was refined (K = 1).
- Error model 6: Individual K and individual g value for each run.
- Error model 8: Individual K and fixed g value. To investigate the effect of only using the scaling parameter, g was set to 0.

The g value is also required in the parameter weighting, which occurs before error model is applied, as weighting scheme applied to the parameters contains g (Equation 1.27). The choice of g at this stage can affect the final value calculated for g in the error model; the treatment of the g parameter for the differing error models is summarised in Table 3.20. In the reference model, the value for g was refined to convergence, with the parameter and error model staged being repeated until g remained the same (see Section 3.2.2). When applying models 0 and 8, where g in the error model was set to 0, the value for g for the parameter weighting was also input as 0. For error model 6, the value of g for the parameter weighting was chosen to be 0.0133, the same as the value of g calculated for the reference model. This was so the difference to the reference model would be due to the application of the error model. For error model 1, the g value was refined to convergence in the same way as in the reference processing (g = 0.0119). All other scaling choices remain unchanged from the reference model in order to be able to attribute the differences between the multipole models produced solely to the choice of error model.

Error model	g parameter treatment
0	Set to 0
1	0.0119 (refined to convergence)
Reference (5)	0.0133 (refined to convergence)
6	Set to 0.0133
8	Set to 0

Table 3.20 g parameter treatment for each error model refinement.

The scaled values for the reflections produced when using the four error models were compared to those calculated for the reference model (default error model). A graphical comparison of the intensity and standard uncertainty values for error model 0 (where no model is applied) shows little difference for the value of the intensity (Figure 3.23). In the graph of standard uncertainties, standard uncertainties for error model 0 are smaller than the reference model for all but the smallest values.

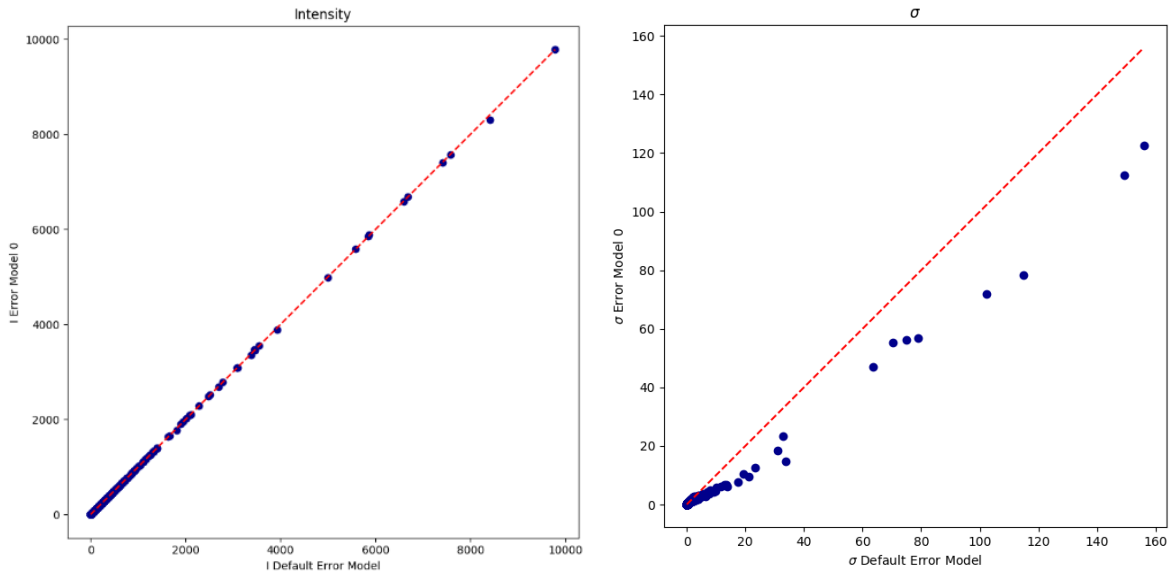


Figure 3.23 Intensity and standard uncertainty comparison for values for reference model and no error model (0). The red dashed line indicates x=y on the plot.

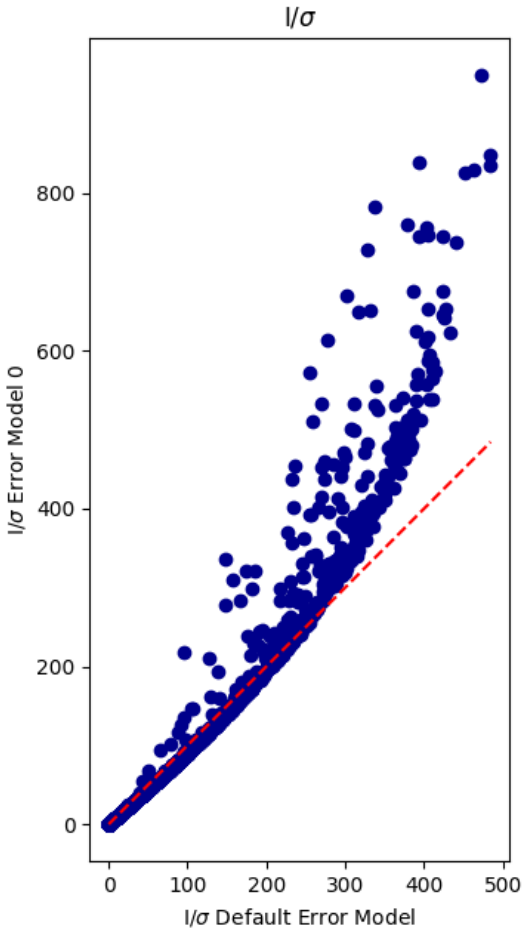


Figure 3.24 I/σ_I plotted for reflections for the default error model, as applied in the reference refinement, vs no error model (0) applied. The red dashed line indicates x=y on the plot.

The difference in the significance of the values for error model 0 and the reference model is more pronounced at larger values of I/σ_I (Figure 3.24). This is due to the smaller standard uncertainties, caused by the removal of the intensity contribution to the standard uncertainty. Despite the increase in the significance of some reflections, there are many

where the significance is actually lower when no error model is applied, which is counter to the published work on the significance of Bragg reflections[79–81] and the cause of which is not immediately obvious. A similar feature was observed in a graph of the comparison of the significance of Bragg reflections to raw reflections by Jørgensen *et al.*,[81] using data scaled with *SORTAV* and *SADABS* error models 0 and 5 (Figure 1.16), but this feature was not commented on.

The reduction in the σ_I values (and hence reduction in I/σ_I) for some standard uncertainties when no error applied is attributed to the scale factor, K, which is calculated individually for each run in the reference refinement. A graphical comparison of the significance of the unmerged reflections (Figure 3.25), where points are coloured by the scale factor applied in reference refinement, shows a trend in the K parameter vs the value of I/σ_I for the reflection. For a small number of runs in the reference refinement, the value for K is calculated as less than 1. As the sum of the uncorrected standard uncertainty and the intensity corrected term is multiplied by K, those reflections with lower intensities and a $K < 1$ return a value of I/σ_I that is lower than when no error model is applied. The trend reverses when $I/\sigma_I < 0$ (Figure 3.26).

When reflections are merged, the difference in K and values for standard uncertainty will result in some standard uncertainties for merged reflections being lower than when no error model is applied. A similar distribution is observed when error model 8 is applied (only K parameter scaled) as shown in Figure 3.27. For the other error models 1 and 6, the difference in the significance is not as great (Figure 3.28).

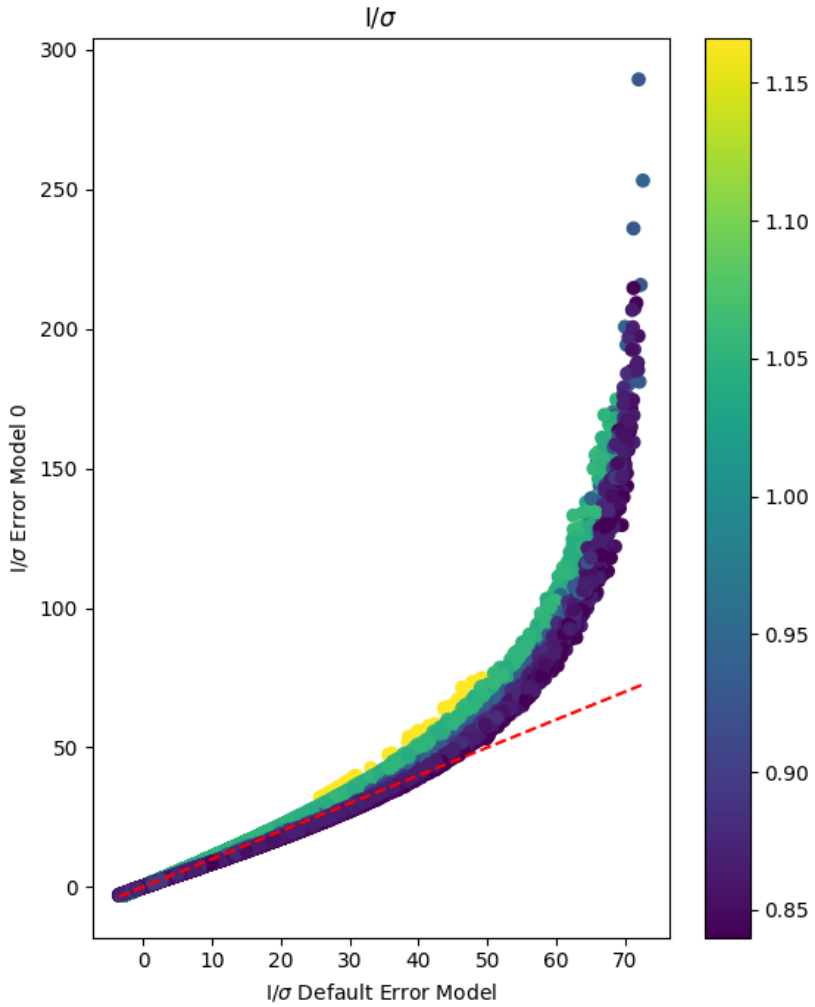


Figure 3.25 I/σ of xd.hkl for default error model (reference refinement) vs error model 0 of unmerged data. The colourbar shows the range in the scale factor, K, applied to the standard uncertainty within the reflections run. The red dashed line indicates x=y on the plot.

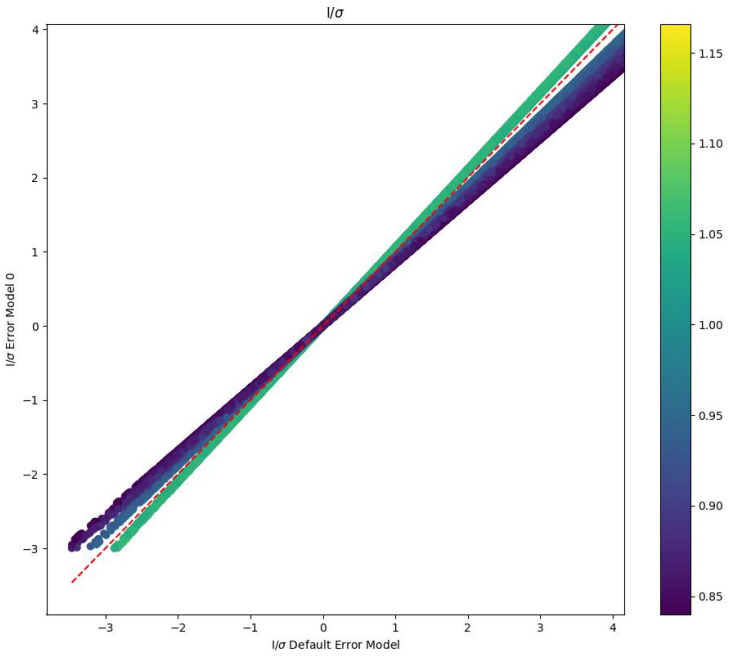


Figure 3.26 Close up of low significance reflections for default error model (reference refinement) vs error model 0 of unmerged data. For significances above zero, reflections where the applied scale factor K, is less than 1 have a larger significance when compared to that of no error model applied.

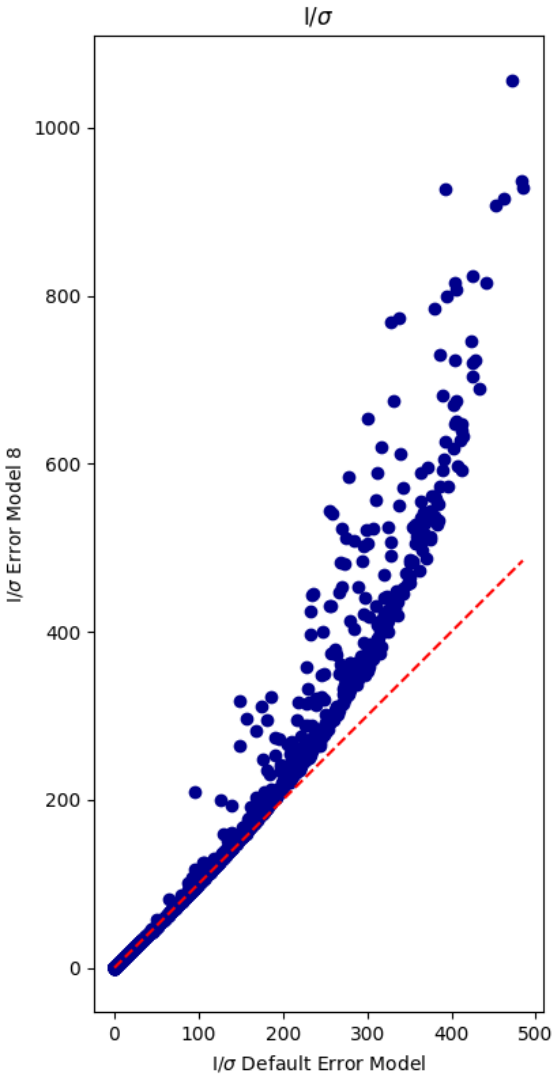


Figure 3.27 I/sig of xd.hkl for default error model (reference refinement) vs error model 8.

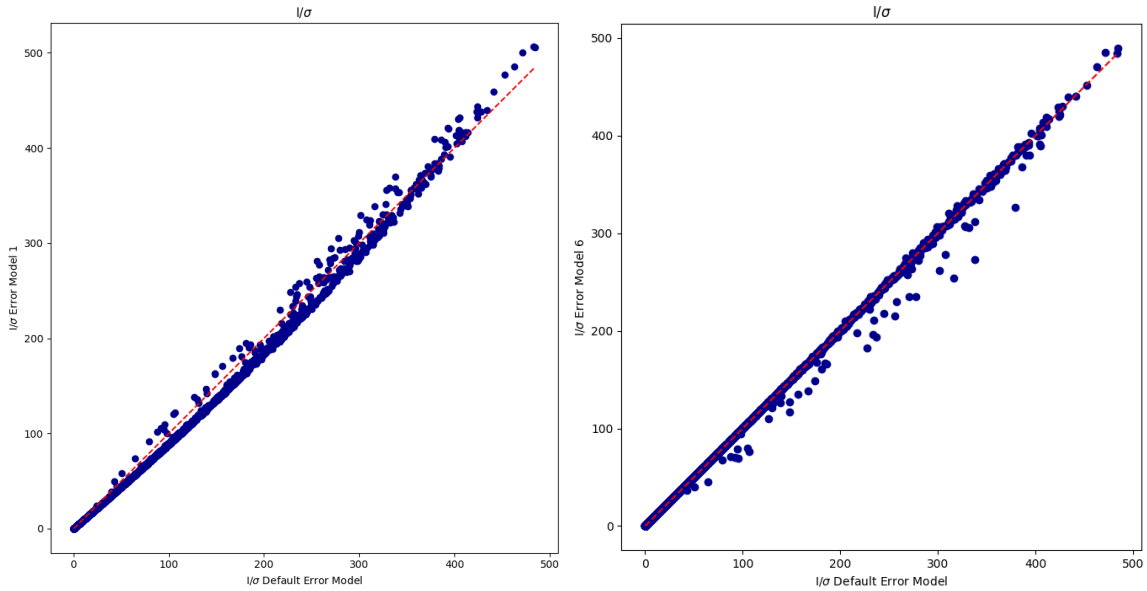


Figure 3.28 Statistical significance of intensity values for the default error model (reference refinement) versus error models 1 and 6.

Multipolar refinement was performed using the same refinement strategy as the reference refinement and the same model starting points (*xd.mas*, *xd.inp*) but the different respective

.hkl files, as produced in *SADABS*. This was deemed acceptable as no difference in the unit cell values measured using the different *.hkl* files was observed. The initial refinement of atomic positions and ADPs within the multipolar refinement strategy would also account for any differences in the positions of ADPs due to the difference in reflection significance.

	Reference	Error model 0	Error model 1	Error model 6	Error model 8
# Reflections	4029	3965	3966	4032	4020
R_2 (%)	1.55	1.48	1.50	1.59	1.50
R_1 (%)	1.48	1.47	1.44	1.48	1.50
wR_2 (%)	1.71	1.40	1.62	1.71	1.42
GooF	1.9668	1.9610	1.7966	1.9511	2.1523
$d^f(0)$	2.4853	2.4915	2.4871	2.4830	2.4910
e_{net} (e)	-0.0320	-0.0507	-0.0358	-0.0353	-0.0490
e_{gross} (e)	1.6545	1.5815	1.6047	1.6775	1.6155
e hole ($eÅ^{-3}$)	-0.174	-0.197	-0.180	-0.168	-0.194
e peak ($eÅ^{-3}$)	0.204	0.184	0.199	0.209	0.187

Table 3.21 A table of refinement statistics for multipole models created using data processed with different error models.

Table 3.21 lists the refinement statistics for the final multipolar models created when modelling against the different *.hkl* files. As an I/σ_I cut-off of reflections < 3 was applied to the refinement, the difference in significance of the reflections due to the error models applied in scaling resulted in a different number of reflections being above this threshold. Error models which included a K parameter for the standard uncertainty (reference, 6 and 8) would have a higher proportion of lower intensity reflections with smaller standard uncertainties (due to K parameters being < 1 for some runs) and hence more reflections are above the I/σ_I threshold.

The range of residual electron density was comparable across all refinements (~ 0.38 $eÅ^{-3}$) as was the shape of the fractal dimensionality plots (Figure 10.15). In a comparison of the flatness of the residual electron density, the values of $d^f(0)$, e_{net} and e_{gross} give contradicting trends making assessment of any improvements to the distribution difficult ($d^f(0)$; 0 > 8 > 1 > reference > 6, e_{net}; reference > 6> 1> 8 > 0 and e_{gross}; 0 > 1 > 8 > reference > 6). $d^f(0)$ and e_{gross} suggest that models, 0 then 8 and 1 have a more featureless residual electron density than the reference refinement model and 6, whereas e_{net} is the smallest for default model suggesting a smaller amount of residual electron density present.

The R_2 factors for error models 0, 1 and 8 are all smaller than those of model 6 and the reference refinement model. The lowest wR_2 values are found for models that would have smaller statistical errors at large values (models 0 and 8) as the intensity contribution to the standard uncertainties is not included. However, the smaller standard uncertainties

calculated for the refinement resulted in a greater range of structure factor residuals, as observed in the normal probability plots of the refinements.

All normal probability plots reported a non-normal distribution of errors (Figure 10.16). Even though some error models return better values, if standard uncertainties are underestimated, removing the error model exacerbates the underestimation in most cases. As there is little difference in the I/σ_I values for error model 6 when compared to the reference refinement, no real improvement in the model was reported – despite this being recommended as a better error model to use for charge density refinement. There are no statistical differences in either the multipolar parameters, net atomic charges or BCPs when compared to the reference model (Section 10.1.8).

All error models have their pros and cons when it comes to the fit of the model to the data e.g. error model 0 has the best refinement statistics, but a larger distribution of residuals in both electron density and structure factors. As the multipolar (all values of S_d below 1) and other model parameters are not statistically different from the default error model, the overall effect of changing the error model in the case of this data is minimal.

3.5.2 Application of overload and numerical absorption corrections

A correction for overloaded reflections and a numerical absorption correction are applied in the reference refinement. Both of these techniques are recommended for use with charge density refinements. Information from fast experimental data collections (or 'fast scans') are used in the scaling process, to provide intensities for reflections which are overloaded in longer data collections. To apply a numerical absorption correction, the faces of the crystal must be located in order to determine the shape of the crystal, which requires a crystal video to be recorded.

Data were processed in *SADABS* with a combination of:

- No overload correction
- No numerical absorption correction
- No overload or numerical absorption correction

The default error model was applied in each case, with a value for g of 0.133 obtained for all models. For the processing with no fast scan or numerical absorption correction, the additional μr absorption correction was applied as 0.2 (the default value suggested by the program), instead of 0.015 as was used in reference data processing (when the size of the

crystal was taken into account). This was to mimic a refinement performed in the absence of this information i.e. where no fast scan or crystal video had been collected.

Multipolar refinement was performed using the same starting point as the reference data processing. Many different parameter lists were trialled, but the multipolar parameters were only found to converge for the models where the overload correction was not applied when the I/σ_I cut-off was 2. Hence for all refinements the same parameter list for the reference processing with the I/σ_I cut-off reduced was used. The completed multipole models were thus compared to $I/\sigma_I < 2$ multipolar refinement (from Section 3.4.3).

Property	$I/\sigma_I < 2$	No numerical absorption correction	No overload correction	No overload or numerical absorption correction
# Reflections	4201	4201	4193	4193
R_2 (%)	1.54	1.56	1.49	1.49
R_1 (%)	1.58	1.59	1.61	1.61
wR_2 (%)	1.72	1.73	1.73	1.73
GooF	1.9229	1.9443	1.9428	1.9428
$d^f(0)$	2.4909	2.4907	2.5051	2.5050
e_{net} (e)	-0.0280	-0.0333	0.0205	0.0207
e_{gross} (e)	1.7560	1.7626	1.6234	1.6199
e hole ($eÅ^{-3}$)	-0.181	-0.178	-0.185	-0.185
e peak ($eÅ^{-3}$)	0.221	0.222	0.221	0.221

Table 3.22 Refinement statistics for different intensity corrections.

Table 3.22 shows the refinement statistics for the three models and the statistics from a model where both corrections were included ($I/\sigma_I < 2$) for comparison. It is evident that the application of the absorption correction makes little difference to the final results (especially when no fast scan is used as the refinement statistics are almost the same). Interestingly when the fast scan is removed the R_2 factor is improved. The fractal dimensionality analysis plots all exhibit the same form, showing a shoulder in the positive residuals indicating features present in the residual electron density (Figure 10.20). Normal probability plots produced for each refinement were also approximately the same (Figure 10.19) – therefore the largest outlier reflections are not affected by the lack of fast scan correction (so the low angle outliers are not due to replacement of data from the fast scan).

There is no statistical difference in multipolar parameters between the reference model and the three different multipole models listed above. All multipolar parameters are within 1 standard uncertainty of those of the reference model.

As the shape of the crystal was quite regular, little difference in the path length of reflections would be expected, which may be why the absorption correction makes only a slight difference to the final model. In the reference refinement, only 9 reflections identified as overloaded and replaced with data from the fast scan, which could explain why there was only a small difference when this was not applied. For a crystal with more overloaded reflections, or with a more rectangular shape, this may not be the case (see Section 6.1 for discussion of the application of a numerical absorption correction to a rectangular crystal).

3.6 Varying integration options

The integration process also offers a number of different algorithms when processing the data which are investigated below with respect to the fit of the model to the data. As results for this section were very similar to those in previous sections, only a brief summary of the investigations is given.

3.6.1 Simple summation

The default integration algorithm in *APEX3* is to measure the intensity of the more significant reflections ($I/\sigma_I > 8$) with the simple summation algorithm, while less significant reflections are measured by the application of a reflection profile (calculated using the shape of those reflections with an $I/\sigma_I > 10$). The profile fitting of the less significant reflections can be turned off in either the *SAINT* program itself or the within the *APEX3* GUI (which operates *SAINT*). While the profile fitting algorithm has been shown to improve the statistics for data with many weak reflections for CCDs,[74] this has not yet been investigated using newer detector technologies. The profile fitting algorithm can sometimes struggle with data collected on a HPAD due to the small point spread function of the detector resulting in reflections being measured over only a few pixels – the algorithm should therefore also be tested with data collected on a CPAD.

To investigate the effect using only the simple summation algorithm on the final refinement, all other options (best-plane background, resolution cut-off for integration) were kept the same as the reference data processing (where the default algorithm was used). Data were scaled in *SADABS* using a correction for overloaded reflections, numerical absorption correction and the default error model. The refined g value calculated for the model was smaller than that of the reference model (g= 0.0114), which leads to smaller σ_I values due to the form of the error model applied (Equation 1.28). As with error model refinements, there is little difference in the values of the intensity between the two sets of reflections. The

standard uncertainties for reflections are in general smaller for the simple summation integration, resulting in an increased significance of the reflection data.

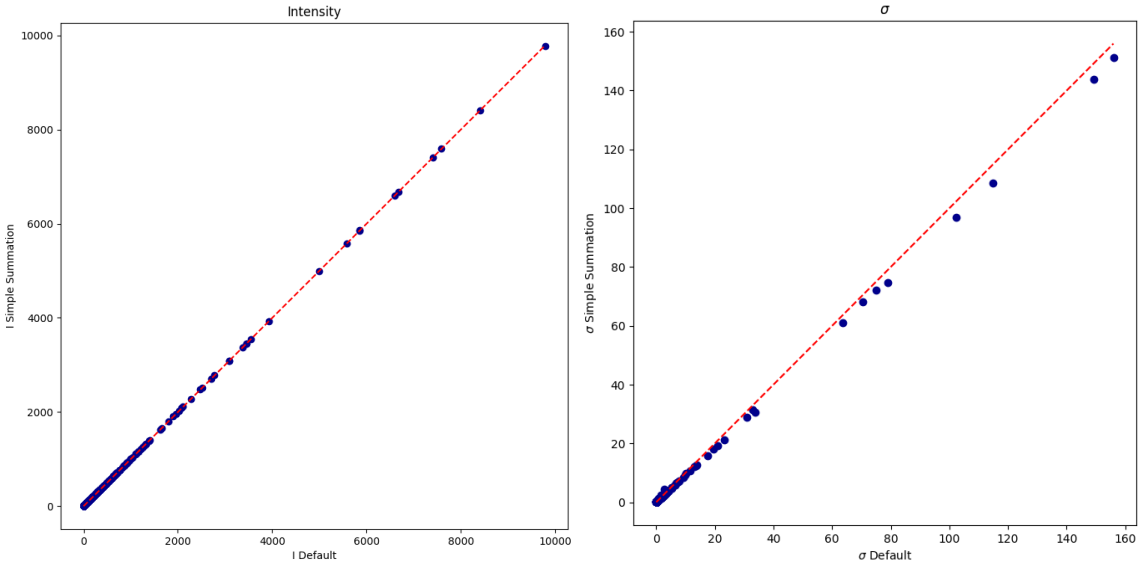


Figure 3.29 Intensity and σ_I comparison for simple summation and default integration algorithm.

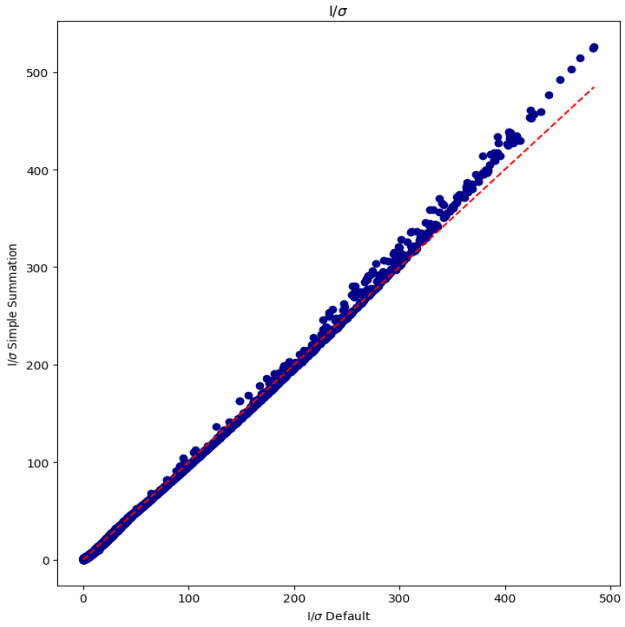


Figure 3.30 I/σ_I for simple summation and default integration algorithm.

Multipole refinements were undertaken using the same parameter list as the reference refinement. The same starting point as the reference refinement was used, with the *.hkl* file being taken from the simple summation processing. Table 3.23 shows the difference in refinement statistics for a simple summation and the default integration algorithm. The values for all parameters are comparable, with an improved GooF being reported for the simple summation refinement. However, there is an increased amount of residual electron density reported from the simple summation algorithm, which could indicate more noise

within the refinement. Normal probability and fractal dimensionality plots were of a very similar shape to the reference model.

Property	Reference	Simple summation
# Reflections	4021	3874
R_2 (%)	1.55	1.58
R_1 (%)	1.48	1.53
wR_2 (%)	1.71	1.61
GooF	1.9668	1.9109
$d^f(0)$	2.4853	2.4896
e_{net} (e)	-0.0320	-0.0266
e_{gross} (e)	1.6545	1.7792
e hole ($eÅ^{-3}$)	-0.174	-0.196
e peak ($eÅ^{-3}$)	0.204	0.213

Table 3.23 Refinement statistics for simple summation and default integration algorithm.

There is no statistical difference in multipolar parameters between the reference model and the simple summation models. All multipolar parameters are within 1 standard uncertainty of those of the default model. Properties calculated from multipolar parameters are also not statistically different. Therefore, the use of only simple summation algorithm to integrate the data does not appear to give improved results in the multipole model in this case.

3.6.2 Resolution limit for integration

A resolution cut-off beyond which no reflections are integrated can be applied to the integration in *SAINT*. This is implemented to remove reflections which may be weak or areas of the diffraction image that are beyond the diffraction range of the crystal (where no diffraction pattern is present). The reference refinement (0.48 Å) is compared against data integrated with a 0.45 Å resolution limit.

The data for the 0.45 Å resolution integration were processed as in the reference refinement. The reflections were scaled in *SADABS* using a fast scan, numerical absorption correction and the default error model, where the value for g was optimised to 0.0135. This is a very similar value to the reference refinement. Multipole refinement took place using the same starting points as the reference model with the same refinement strategy.

The refinement statistics (Table 3.24) show that there is an increased amount of residual electron density within the model when data with a higher resolution cut-off is used. This may be due to an increased amount of noise due to weak or poorly measured reflections at high angles. This is also mirrored in the residual electron density plot (Figure 3.31), where there is a greater amount of residual density present when compared to the reference model (Figure 3.6). Henn and Meindl showed that increasing the resolution limit and

including more data resulted in a larger value for $d^f(0)$.[37] This is reflected in the refinement statistics.

Property	Reference (0.48 Å cut-off)	0.45 Å cut-off
# Reflections	4021	4565
R_2 (%)	1.55	1.62
R_1 (%)	1.48	1.72
wR_2 (%)	1.71	1.79
GooF	1.9668	1.9187
$d^f(0)$	2.4853	2.5087
e_{net} (e)	-0.0320	-0.0587
e_{gross} (e)	1.6545	1.9107
e hole ($eÅ^{-3}$)	-0.174	-0.182
e peak ($eÅ^{-3}$)	0.204	0.211

Table 3.24 Refinement statistics for reference and refinement with data integrated with a higher resolution cut-off (0.45 Å).

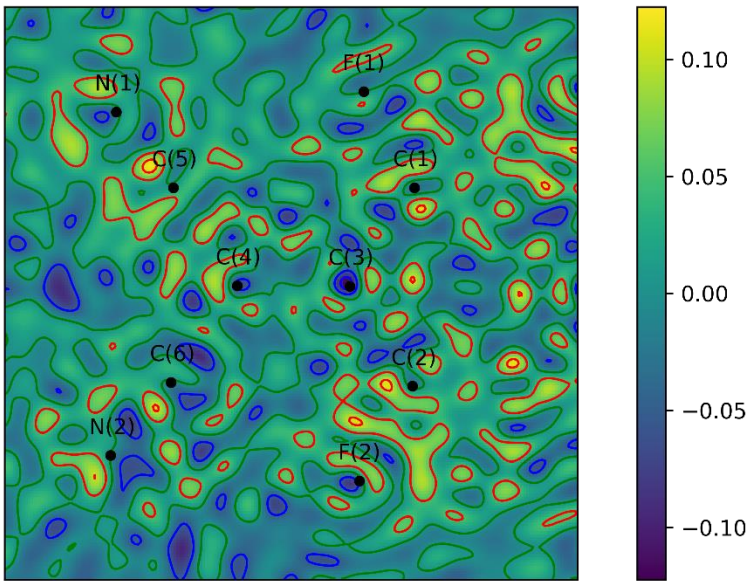


Figure 3.31 Residual electron density plot for 0.45 Å resolution limit refinement. Contour values are drawn at levels of ± 0.05 $eÅ^{-3}$, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green.

There is no statistical difference between the multipolar parameters for either model, with all values lower than 1 σ_I. There is also no statistical difference for any parameters calculated for the model. Statistically, applying a resolution based cut-off should not bias the residuals of the refinement (as occurs with intensity or significance cut-offs),[83] therefore integrating to a lower resolution may be thought of as reasonable, especially when it results in improved refinement statistics. Charge density usually performed with a resolution beyond 0.5 Å; a decrease of the resolution limit to 0.5 Å may improve refinement statistics, however, sacrificing reflections with greater contribution from core atoms could result in less accurate atomic positions.

3.6.3 Phosphor efficiency

The phosphor efficiency is a correction which is applied to the intensities and standard deviations during the integration stage. This is applied to detectors which contain a phosphor and corrects for X-ray photons lost when they are converted to visible light. The default value for the Photon II detector at Newcastle University is 0.68%, however, it was felt that it was important to ensure this was the best value of the correction.

The phosphor efficiency correction is a multiplicative correction applied to both the intensities and standard uncertainties that is dependent upon the position the on the detector at which the reflection was measured. Figure 3.32 shows the form of the correction for one run of the data integrated using the default phosphor efficiency versus an integration where the correction was turned off. This shows that the value applied to the intensity and standard uncertainty depends on the position of the reflection on the detector. Reflections which were measured further away from the centre of the incident beam have a larger correction applied to them. The correction applied to the intensities and standard uncertainties is linear therefore the significance of the reflection values does not change before scaling is applied (when comparing values as reported in the *.raw* file).

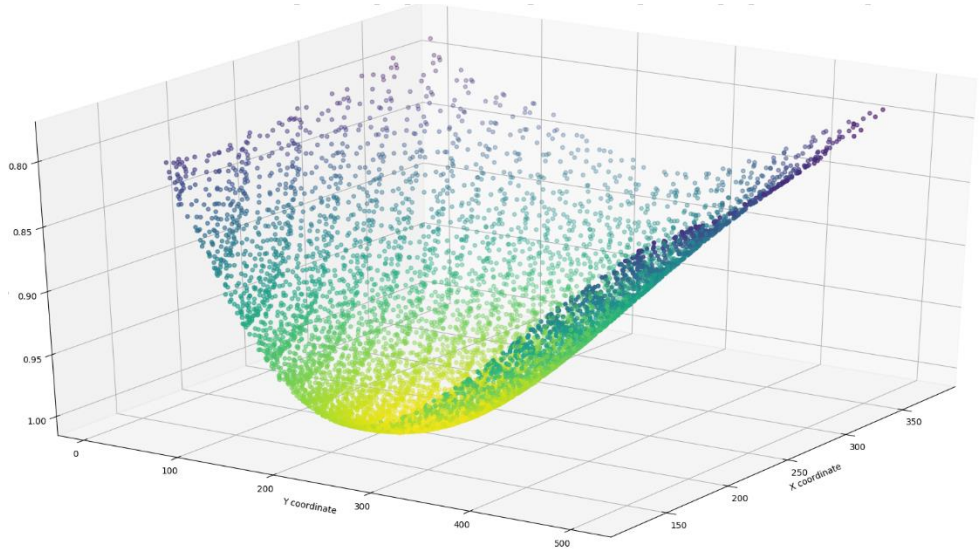


Figure 3.32 A 3D plot of the detector position the reflection was measured at (X and Y) with respect to the difference in I for a phosphor efficiency of 0.68 Å and the correction turned off (1.0).

To investigate the optimal value for the phosphor efficiency, data were integrated with varying values for the efficiency. These data were then scaled in *SADABS* using a fast scan, numerical absorption correction and the default error model. The g value was approximated as the experimental error for the refinement; the value for the phosphor efficiency that

produced the smallest g value was therefore taken as the optimal value. Figure 3.33 shows a plot of the distribution of g values reported for phosphor efficiencies of between 67% to 75%. The g value reduced to a minimum value with an increase of the phosphor efficiency value to 70.5% (g = 0.0129) before rising again. It should be noted that g is only reported to 4 decimal places in *SADABS*, which is why there appear to be plateaus in places on the graph. Beyond the range reported, the g value increased further.

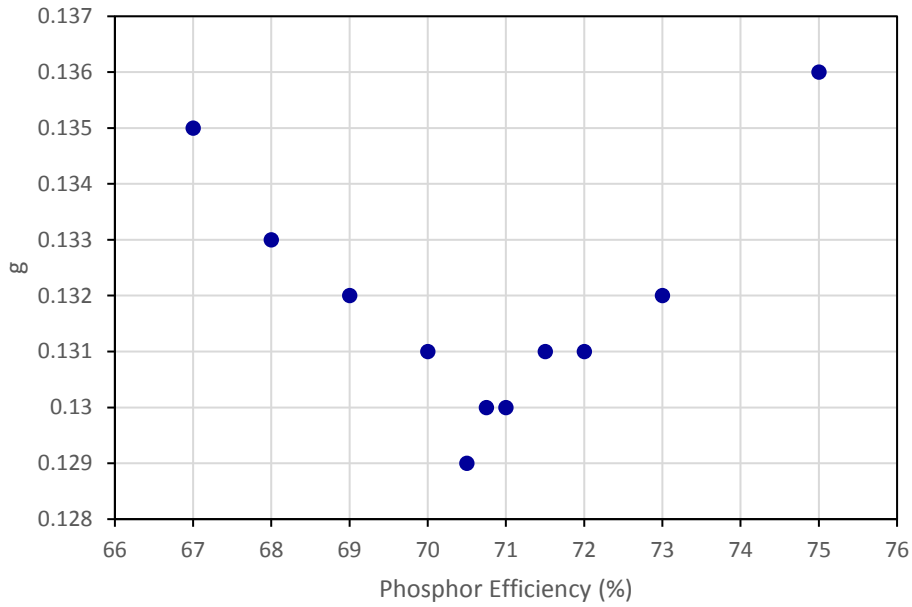


Figure 3.33 A plot of converged g value against phosphor efficiency correction applied.

The reference model was compared to that of data integrated with a phosphor efficiency of 70.5%. The data was otherwise processed in the same manner as the reference refinement. This resulted in a multipolar model with a very similar refinement statistics (Table 3.25). The value for the R factors does improve when the 70.5% correction is applied, although the value for the GooF increases. The range of residual density is comparable in both models – 70.5% model has slightly improved values for $d^f(0)$ and e_{gross}, which could imply that this is a better model.

	Reference (68%)	70.5%
# Reflections	4021	4027
R_2	1.55	1.53
R_1	1.48	1.48
wR_2	1.71	1.70
GooF	1.9668	1.9724
$d^f(0)$	2.4853	2.4867
e_{net} (e)	-0.0320	-0.0320
e_{gross} (e)	1.6545	1.6429
e hole ($eÅ^{-3}$)	-0.174	-0.175
e peak ($eÅ^{-3}$)	0.204	0.203

Table 3.25 Refinement statistics for reference refinement (68%) and an optimised phosphor correction of 70.5%.

Again, there is no statistical difference between the multipolar parameters for either model, with all values lower than 1 σ_l. There is also no statistical difference between the properties calculated from the model. The reduced error model as calculated in *SADABS* and the improved refinement statistics suggest that the optimal phosphor efficiency correction for the detector used in this investigation is 70.5%.

3.7 An optimal refinement strategy

Of the different data processing strategies that have been trialled, many make very slight differences to the overall multipolar parameters and the properties which can be calculated from the multipolar model. Some models have a better fit to the data, indicated by improved refinement statistics, but the aim of all crystallographic refinements is the production of a chemically sensible model – which should be prioritised over a reduction in R factors and residual electron density.

The application of a *SHELXL* weighting scheme returns an approximately normal distribution of structure factor residuals. In the reference refinement the shape of the normal probability plot is the same as the underestimated standard uncertainty plot in Henn and Meindl,[40] implying the presence of this systematic error within the investigated data. The underestimation of standard uncertainties for reflections in charge density is an accepted source of systematic error, therefore, the weighting of the model to correct for this is a reasonable action. As residual electron density and fractal dimensionality analysis showed little increase in the features when compared to the unweighted multipolar model and there was no statistical difference in the multipole parameters or parameters calculated from the model, it is suggested that in this case the *SHELXL* weighting scheme is an accepted addition to the multipolar model.

It is also of interest that not including a fast scan to address the problem of overloaded reflections resulted in a better overall fit. The reflections taken from the fast scan were measured for a shorter exposure time, and thus would have much larger standard deviations and be measured less accurately than other reflections. For this particular dataset the number of overloads replaced by the fast scan was small (9), in datasets with more overloads this may not be the case.

The data processing options investigated in the chapter appeared to make little difference to the final model when compared to the reference model in isolation. This may not be the

case if refinement processing options were combined. While these processing options may not make statistical difference, it may be preferable to do the data in a 'more correct' manner, e.g. minimising the experimental error by applying a phosphor correction of 70.5% or using the suggested error model 6 for scaling data.

It is difficult to draw conclusions on differences in properties such as the net charge as the values were accompanied by large respective errors and many parameters were not statistically significant. Furthermore, the analysis of statistical difference depends on the standard uncertainties for parameters being evaluated correctly. In the *XD2016* software the calculation of errors on the density and the Laplacian only take into account contributions to the error from multipole populations and not κ or coordinates, meaning errors for these latter parameters could be the same or larger if they were considered. Therefore, Laplacian parameters identified as statistically different may not be when the additional sources of error are considered; only a few Laplacian parameters were identified as statistically different and therefore this should have little effect on the conclusions in this chapter.

These conclusions are all specific to this particular dataset and larger, more complex asymmetric units or different experimental conditions may exhibit a greater variation in these parameters. The fact that the same processing strategy as the reference model did not converge in every case highlights that the refinement strategy was sensitive to the data processing options, even if the final models were not.

3.8 Future work

The investigation into the effect of data processing options on multipolar models described in this chapter is by no means complete. There are many other algorithm choices and cut-offs which could be investigated which are routinely applied in published charge density research. A small selection of these are described below:

- The effect of algorithm or cut-off was generally investigated in isolation (with no other changes in parameters). The application of a combination of different processing choices as discussed above may result in greater deviation from the reference processing.
- Different programs exist for the various stages of the refinement and are used within the field. There are a number of different integration programs which all operate in different ways, such as *CrysAlisPro*[144] and *Xia2*.[145] Integration of data with different programs may produce a greater difference in the reflection intensities or standard

uncertainties than seen in the investigations so far, which may have a greater effect on the final multipole model. *SORTAV* is another popular scaling program routinely used in charge density refinement which could be utilised. Different multipolar refinement programs e.g. *MoPro* could also be investigated.

- There are a number of different databanks from which the atomic orbitals are obtained from. In this study, analysis was restricted to that of the default databank. There are 3 further databanks within *XD2016* which could be investigated.
- The high multiplicity of the reflections measured for charge density is one of the causes of the underestimation of standard uncertainties for refinements, therefore an assessment into the effect of multiplicity on the multipolar refinement with a view to prevent the underestimation of standard uncertainties is vital. In the data presented in this chapter, the multiplicity of reflections decreases with increasing resolution. Some reflections were therefore measured many more times than others, which could create an uneven distribution of standard uncertainties (and may be the cause of the resolution dependence observed in the normal probability plots). An experiment of interest would be to only include the same number of measurements in the scaling for every reflection (i.e. only use 15 measurements of the total number measured for a particular reflection). However, the selection of which reflections to be used if there were more than the allowed number would be a non-trivial process.
- There is further scope for investigation of the integration algorithm used to measure the reflections. The significance threshold of a reflection before the simple summation algorithm is used, along with the significance threshold for use of the reflection in the model profiles (for profile fitting the less significant data) can all be changed within the *APEX* GUI. *APEX3*, along with *CrysAlisPro* and *Xia2,* can also apply the profile fitting algorithm to all reflections. Other integration options, such as the choice of background, could also be interrogated.
- The *SHELXL* weighting scheme contains 6 parameters which can be varied by the user, although only 2 are currently optimised in *CAPOW*. Additional parameters could be included in the weighting optimisation for *CAPOW* which could produce a better fit of the calculated and observed structure factors.

Further investigation is required if charge density models are to continue to be used to calculate quantitative information that is used to compare against other multipolar models, otherwise comparison should be restricted to qualitative analysis only.

Chapter 4. Polymorphism of F_4TCNQ

In the process of growing high-quality single crystals of F_4TCNQ suitable for charge density studies, the compound was found to exhibit polymorphism and a crystal structure with a different packing arrangement, a new polymorph, was discovered. The International Union of Crystallography defines polymorphism as an instance where *"the same chemical compound exhibits different crystal structures"*.[146] While the arrangement of molecules differ in the solid state, polymorphs have the same form in liquid or gaseous states. Polymorphism can occur due to differences in crystallisation or experimental conditions (such as the application of pressure or low temperatures)[147] but in this case both polymorphs of F_4TCNQ were observed to form concomitantly, under the same reaction conditions in the same crystallisation vial.

Inflexible molecules, such as F_4TCNQ, tend to exhibit packing polymorphism where the change in structure is *"directed by intermolecular interactions*".[16] Intermolecular interactions, while often weak, can provide powerful stabilising interactions when summed over the entirety of the molecule. The structure of the material can affect its properties such as solubility, therefore the crystal structure of different polymorphs is extremely important in drug design, where different polymorphs of active pharmaceutical agents have been shown to have varying bioactivity.[148,149]

Comparisons of the multipole models arising from the two polymorphs or forms allows further investigation into the transferability of multipole parameters, in this case, between the same molecule in different structural arrangements. This comparison would not be affected by different chemical entities as both crystals are of the same pure form. Due to the conformational inflexibility of F_4TCNQ, polymorphs of the same molecule should still result in the atoms in molecules being in crystallographically similar environments within the crystal. This enables the direct comparison of multipolar parameters.

Charge density has been shown to provide additional insight into the formation of different polymorphs through the analysis of intermolecular interactions by identifying intermolecular critical points. Munshi and Guru Row performed analysis on structural differences in the

concomitant forms of 3-acetylcoumarin,[16] and described assessing intermolecular interactions using BCPs as superior to basing the analysis on the close contact of atoms in adjacent molecules. Intermolecular BCPs were also investigated in polymorphs of isonicotinamide-oxalic acid and the contribution of short, strong hydrogen bonds were found to be covalent in nature through analysis of Laplacian values.[150]

4.1 Crystal growth

F_4TCNQ was purchased from purchased from Apollo Scientific as a solid with 97% purity and used without further purification (as described in Chapter 3). Recrystallisation of F_4TCNQ to grow high quality single crystals was attempted in several different solvents, including toluene, acetonitrile and dichloromethane (DCM). Crystals of F_4TCNQ grown from saturated solutions of both acetonitrile and DCM were found to be homomolecular, while crystals grown in toluene were formed as a solvate. Crystal formation took place within 24 – 48 hours.

Further investigation showed that slow evaporation of a saturated solution of F_4TCNQ in DCM yields single crystals exhibiting two different morphologies; cubic crystals, which were the known form of F_4TCNQ (polymorph I), alongside octahedral crystals (one of which is shown in Figure 4.1). These octahedral crystals are the same yellow colour as those of polymorph I. Diffraction data for these octahedral crystals were collected and processed; the structural solution and refinement confirming a novel structural motif different to the known form, hence called polymorph II.

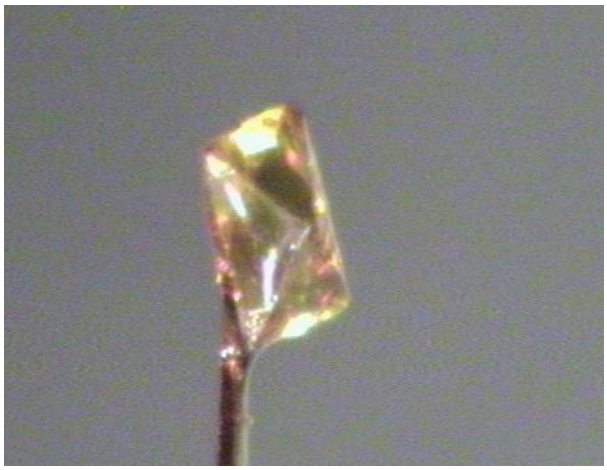


Figure 4.1 Octahedral crystal of F_4TCNQ polymorph II (0.121 x 0.166 x 0.3 mm^3).

4.2 Data collection for polymorph II

The crystal of polymorph II was slowly cooled to 100K at a rate of 1K per minute i.e. the same as for form I and measured using the same instrument as described in the previous

chapter. The collection strategy was calculated to achieve a multiplicity of 12 for 90% of the data up to 0.45 $Å^{-1}$. The experimental run list for the collection of polymorph II is included in Section 10.2.1.

The same combination of data processing options was applied to both polymorphs thereby removing any differences in the final models that would be due to differences in processing. Integration was performed using *SAINT* with 0.48 Å cut-off, default integration algorithm and best-plane background. The data were scaled and merged in *SADABS* using the default error model, a correction for overloaded reflections and a numerical absorption correction based on the faces of the crystal were applied. The space group was identified in *XPREP* and the *IAM* solution and refinement took place in the *Olex2* GUI using *XL* and *XT* respectively.

4.3 Spherical atom model structural comparisons

Initial comparisons of the two structures are possible using the spherical atom model. Table 4.1 shows unit cell parameters for the two polymorphs. Both polymorphs crystallise in the orthorhombic crystal system and centrosymmetric space groups. Form II is slightly denser than form I – suggesting a closer packing arrangement of molecules resulting in a smaller unit cell volume. Z and Z' denote the number of complete molecules in the unit cell and asymmetric unit respectively. The atoms in F_4TCNQ in form II sit on a special position in the unit cell, effectively a horizontal mirror plane, therefore halving the number of atoms in the asymmetric unit relative to form I.

Polymorph	I	II
Space group	*Pbca*	*Pnnm*
Crystal system	Orthorhombic	Orthorhombic
a (Å)	9.1803(3)	7.5140(4)
b (Å)	8.0483(3)	11.6787(6)
c (Å)	14.5540(5)	5.9347(3)
V ($Å^3$)	1075.33(7)	520.79(5)
ρ (g/cm^3)	1.705	1.761
Z	4	2
Z'	0.5	0.25

Table 4.1 Unit cell parameters for polymorphs of II.

Refinement statistics for spherical atom refinements of polymorphs I and II are listed in Table 4.2 and have comparable values for many parameters including the R factors. The R_{int} of the form II with respect to resolution is also similar to form I (Figure 4.2 and Figure 3.3), although the resolution value at which the R_{int} rises above 10% is higher ~ 0.55 Å.

Polymorph	I	II
R_1	3.57%	3.13%
wR_2	11.79%	11.14%
R_{int}	4.14%	4.26%
GooF	1.080	1.066
Total reflections	189770	120718
Independent reflections	5103	2628
Mean I/σ_I	94.6097	131.85
e_{hole} ($eÅ^{-3}$)	-0.323	-0.262
e_{peak} ($eÅ^{-3}$)	0.750	0.773
Weighting scheme	*SHELXL* a = 0.059, b = 0.133	*SHELXL* a = 0.067, b= 0.030

Table 4.2 Refinement statistics for spherical atom refinement of F_4TCNQ polymorph I and II.

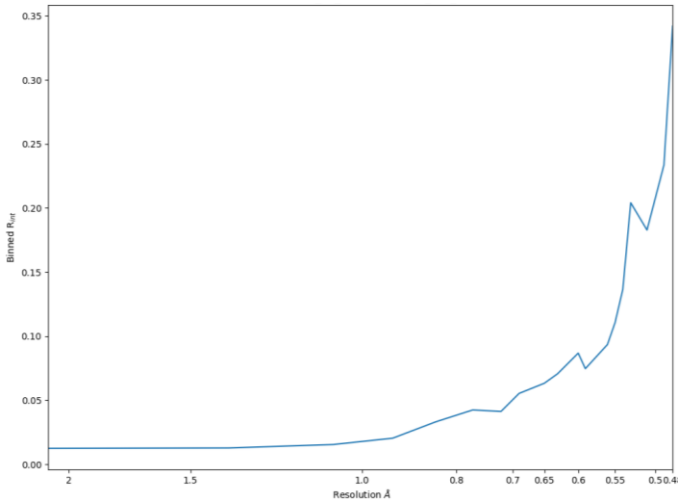


Figure 4.2 Binned R_{int} for reflections included in multipolar refinement for F_4TCNQ polymorph II.

The molecular geometry of both form I and II are extremely similar (with no statistically different bond lengths and angles). This is unsurprising – owing to the conformationally inflexibility of F_4TCNQ due to its almost planar nature, which is a result of the high degree of conjugation within the molecule. The root mean square of the deviation when overlaying one asymmetric unit on top of the other is 0.023 Å, an exceptionally low value which indicates negligible conformational variation.

Despite their similar molecular structures, the packing of the molecules in the two crystal structures is extremely different. The packing of polymorph I can be described in terms of a series of discrete layers, as seen when looking down the crystallographic *b* axis or [0 1 0] direction (Figure 4.3). Molecules are not arranged parallel to the plane, rather they are at an angle of ~ 37 °.

The orientation of the molecules alternates between adjacent layers, as indicated by the arrows in Figure 4.3. The angular arrangement creates a distinctive herringbone pattern of molecules in adjacent layers as illustrated in Figure 4.4, an alternative view of the crystal

structure along the [0 1 1] direction. Adjacent molecules, drawn using the wireframe model, are also arranged in a herringbone formation, but at 90 degrees herringbone chain highlighted in the figure.

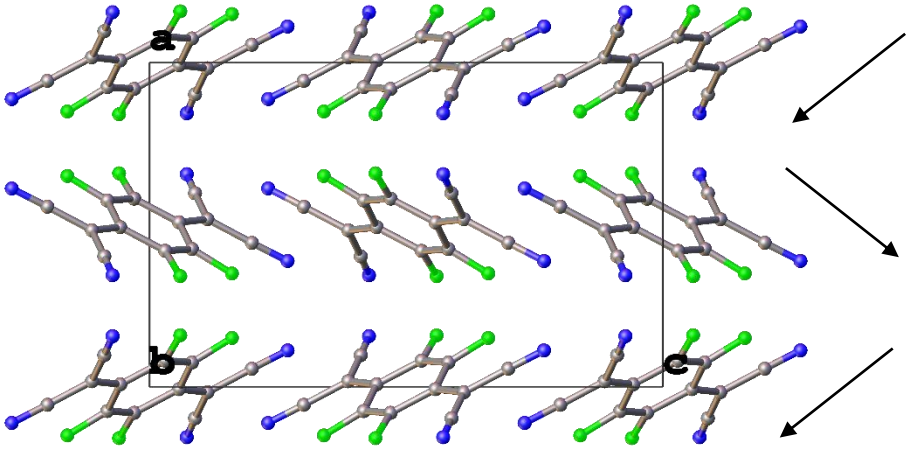


Figure 4.3 View of polymorph I along the [0 1 0] axis. The box represents unit cell.

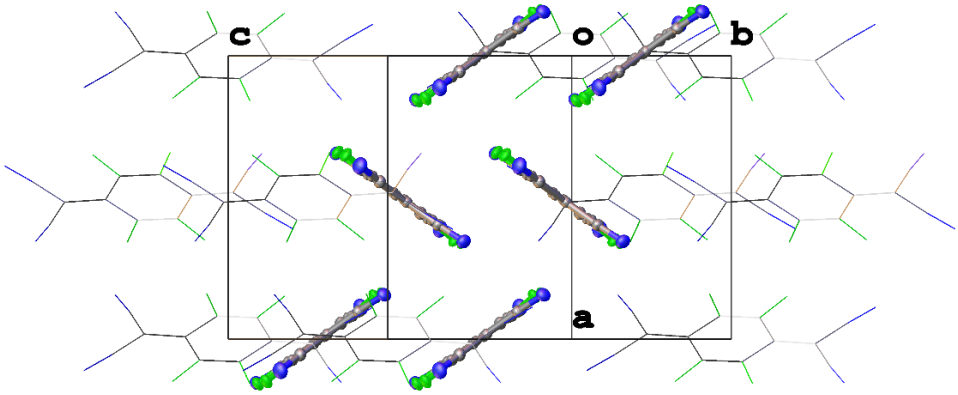


Figure 4.4 View of I along the [0 1 1] direction between layers, showing twisting of molecules in adjacent planes.

Within a layer, as shown along the [1 0 0] direction in Figure 4.5, molecules are arranged in alternating directions, which can also be described as a herringbone pattern.

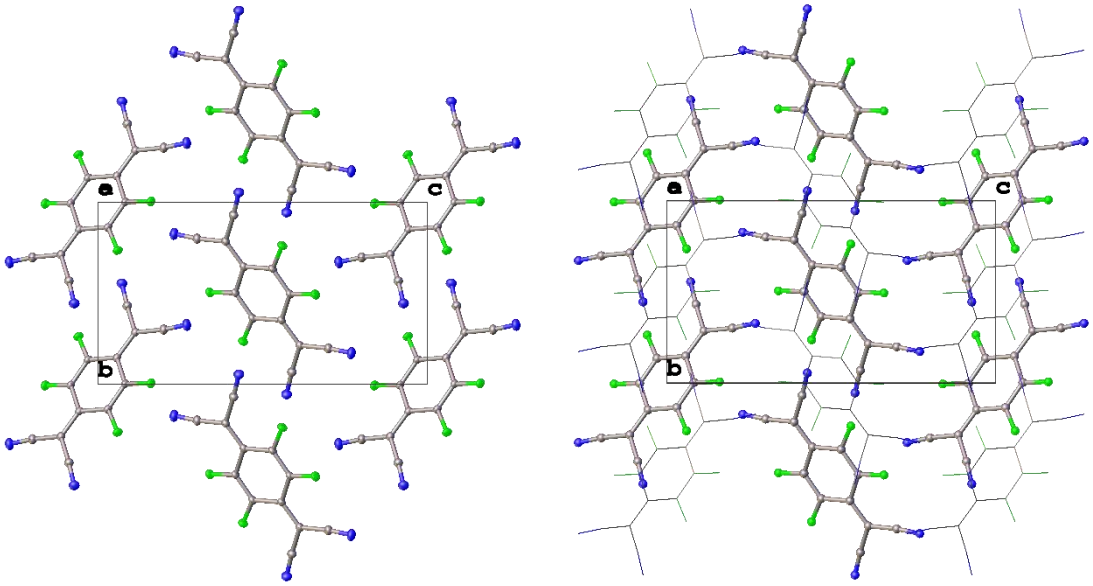


Figure 4.5 A single layer (left) and diagram with adjacent layers visible (right) in [1 0 0] direction of polymorph I.

Polymorph II can also be described as molecules packing together in layers, although in this structure the molecules within each layer are arranged parallel to those in adjacent layers (Figure 4.6).

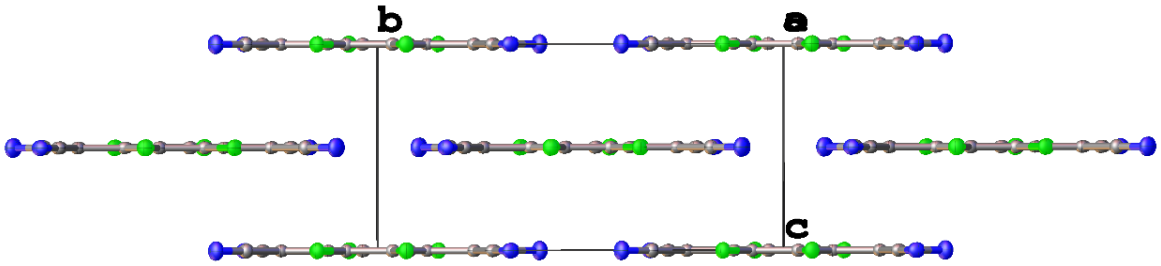


Figure 4.6 View along [1 0 0] axis of polymorph II.

Within a layer, each molecule is related by a crystallographic translation (Figure 4.7). In adjacent layers, the direction of the molecules alternates with respect to the previous layer (Figure 4.8).

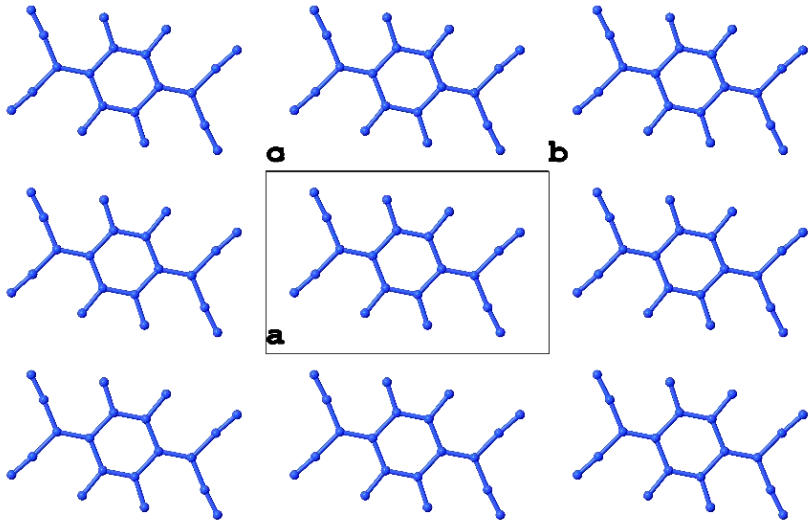


Figure 4.7 View down c axis [0 0 1] of polymorph II, showing only one layer.

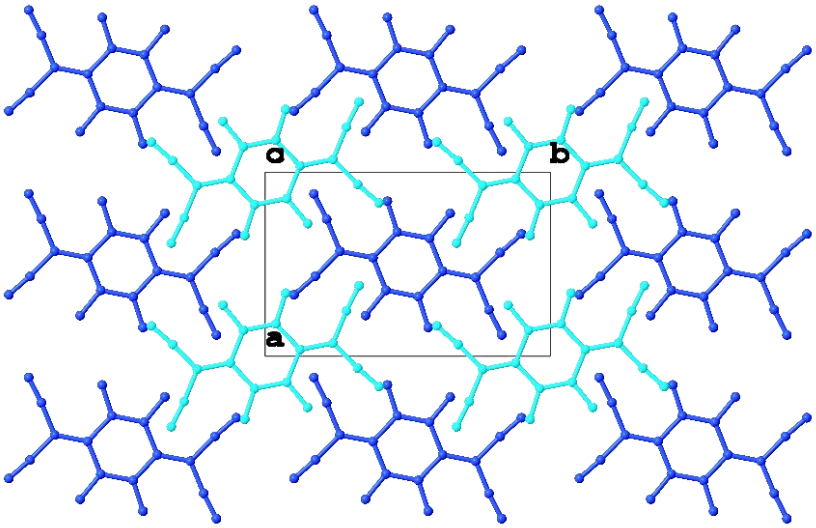


Figure 4.8 View down c axis [0 0 1] of polymorph II. Molecules in different layers are shown in different colours.

4.4 Hirshfeld surfaces

A Hirshfeld surface is a method of partitioning the unit cell that can provide insight into the arrangement of molecules within a crystal – based on a technique developed by Hirshfeld.[141] The 'electron density' within the unit cell is partitioned using the assumptions of the spherical atom model (and does not take into account any multipolar parameters). The Hirshfeld surface encloses the volume within the unit cell where the contribution to the electron distribution of the unit cell from a particular molecule is greater than that of the surrounding molecules.[151,152] A variety of properties can be plotted on the Hirshfeld surface to give additional information about the shape of the surface and its curvature, as well as information about distances between atoms in adjacent molecules. The Hirshfeld surfaces in this work were produced with *CrystalExplorer17*.[153]

One useful property to plot on a Hirshfeld surface is the normalised contact distance, d_{norm} (as can been seen in Figure 4.9). d_{norm} is calculated for all points on the surface as the value of the sum of the distances from the surface to the nearest atom both within the surface and outside the surface, normalised with respect to the van der Waals radii of the corresponding atoms. When d_{norm} is plotted on the surface, areas of the surface where the distance between a molecule inside the surface and a molecule outside the surface is smaller than the total of their two van der Waals radii are shown in red and denote a close intermolecular contact. White represents areas where the distance is equal to the van der Waals radius and blue where the distance is larger.

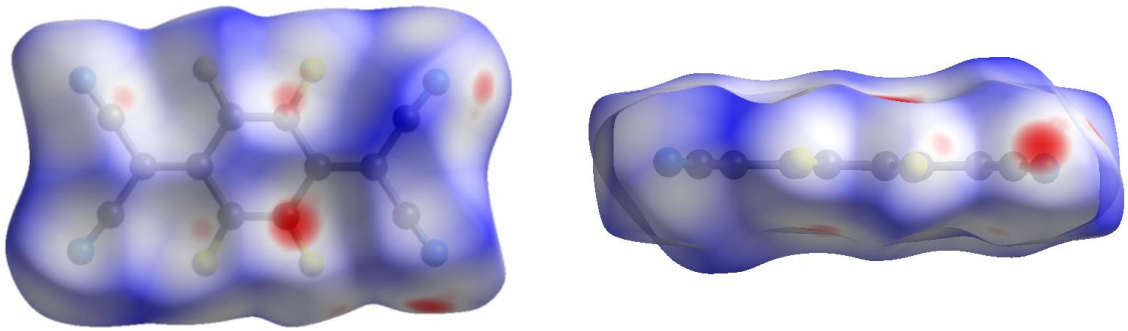


Figure 4.9 Hirshfeld surface for form I, showing d_{norm} values, where red areas indicate where two atoms are closer than the sum of their van der Waals radii.

The contact distance between atoms is important as the strength of many intermolecular interactions are dependent on the distance between the atoms in question (amongst other factors i.e. the angle between atoms). So called 'close contacts' occur when the distance between two atoms is smaller than the sum of their van der Waals radii. This can suggest

likely intermolecular interactions within the structure and thereby assist in the explanation of the crystal packing.

The Hirshfeld surface for polymorph I is illustrated in Figure 4.9. Most of the surface is white or blue, indicating interatomic distances of the van der Waals radius and beyond. There are also a small number of areas where there is a closer than van der Waals contact in red; these areas often correspond to specific interactions. In mostly planar molecules, intermolecular interactions are often described by what part of the molecules interact, either edge-edge, face-face or edge-face. These can be attractive or repulsive depending on which atoms are within close contact range. In form I, the close contacts correspond to a combination of edge-edge and face-edge interactions between molecules within the same layer (Figure 4.10). The close intermolecular contacts as reported by *CrystalExplorer* are between the carbon atoms of the cyano group and between nitrogen and both atoms of the C – F bond.

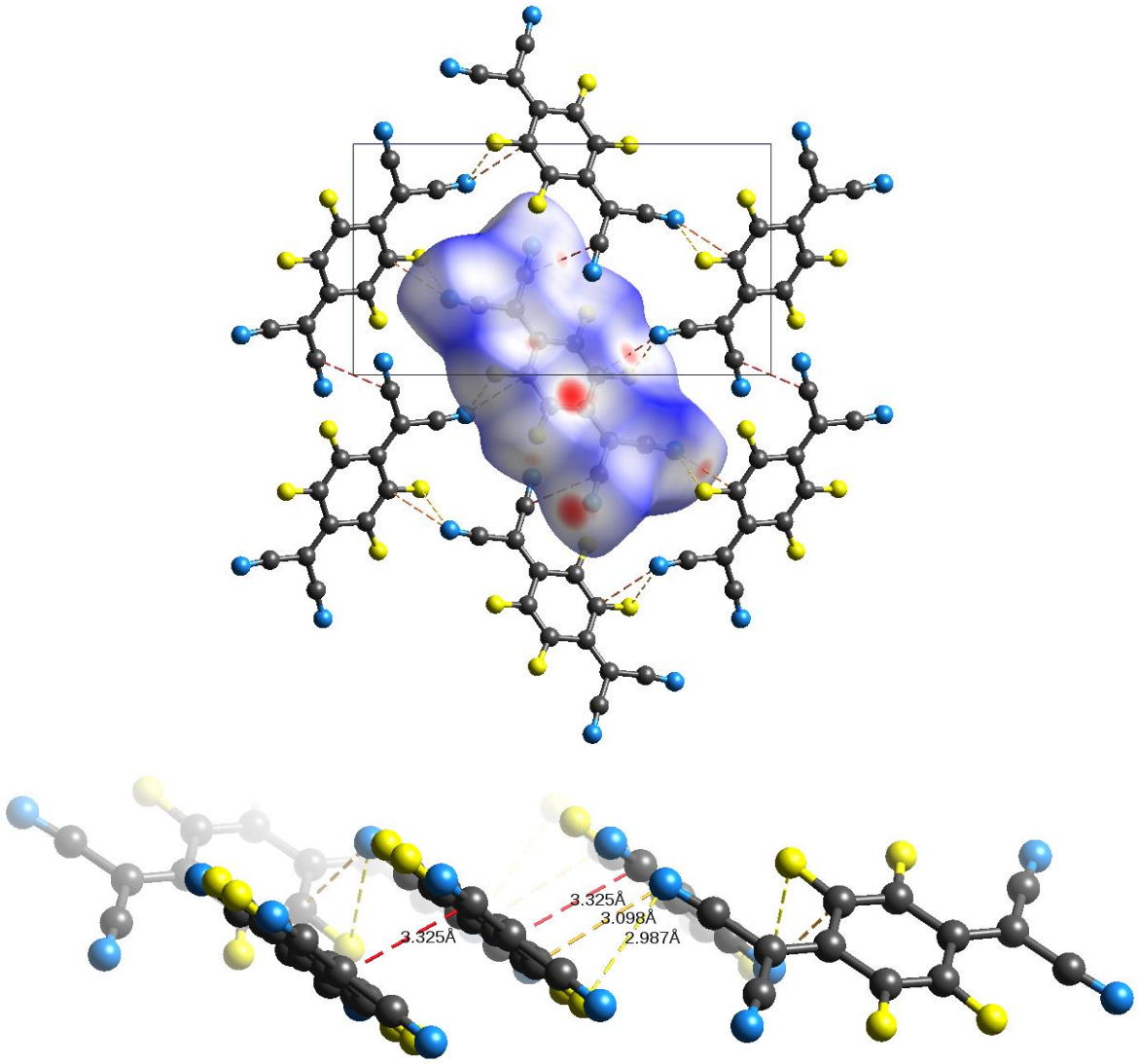


Figure 4.10 Hirshfeld surface for polymorph I, showing molecules within layer which correspond to close contacts. Different colours represent different atomic contacts (red = C···C, orange = C···N, yellow = N···F). Hirshfeld surface removed in second image for clarity.

Between adjacent layers exclusively face-edge interactions are observed (Figure 4.11). These occur between molecules within the herringbone pattern (as described in Figure 4.4). The close contacts observed are between atoms which have intramolecular polar bonds, due to the electronegativity difference of the atoms e.g. C in C – F bond and adjacent F or N in cyano group to F and the adjacent C (which displays a partial positive charge in some charge density structures in Chapter 3). These are likely to be electrostatic interactions.

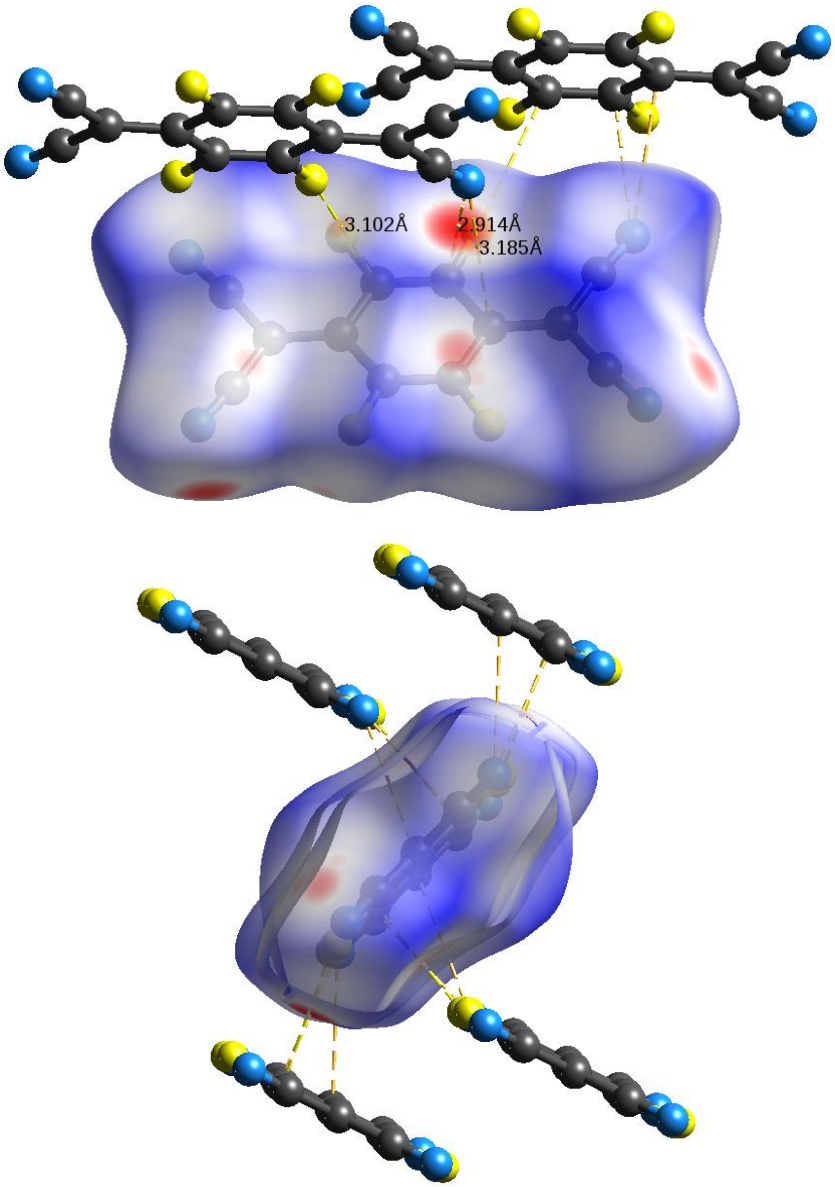


Figure 4.11 Hirshfeld surface for form I, showing molecules in adjacent layers with close contacts to each other. Different colours represent different atomic contacts (orange = C···N, yellow = C···F).

In contrast, when evaluating the Hirshfeld surface of polymorph II (Figure 4.12), many more close contacts between atoms in different layers are observed, with only one distinct contact between atoms in the same layer.

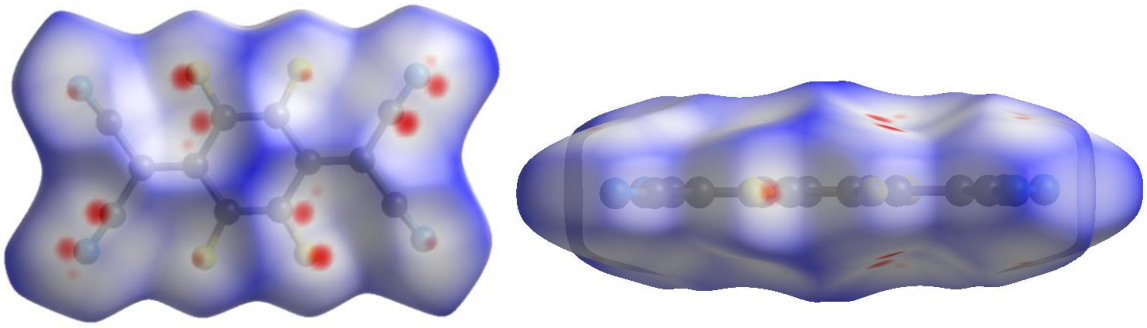


Figure 4.12 Hirshfeld surface for form II showing d_{norm} values, where red areas indicate where two atoms are closer than the sum of their van der Waals radii.

The twisting of molecules in adjacent layers allows close contact between molecules (above and below) in a face – face fashion (Figure 4.13).

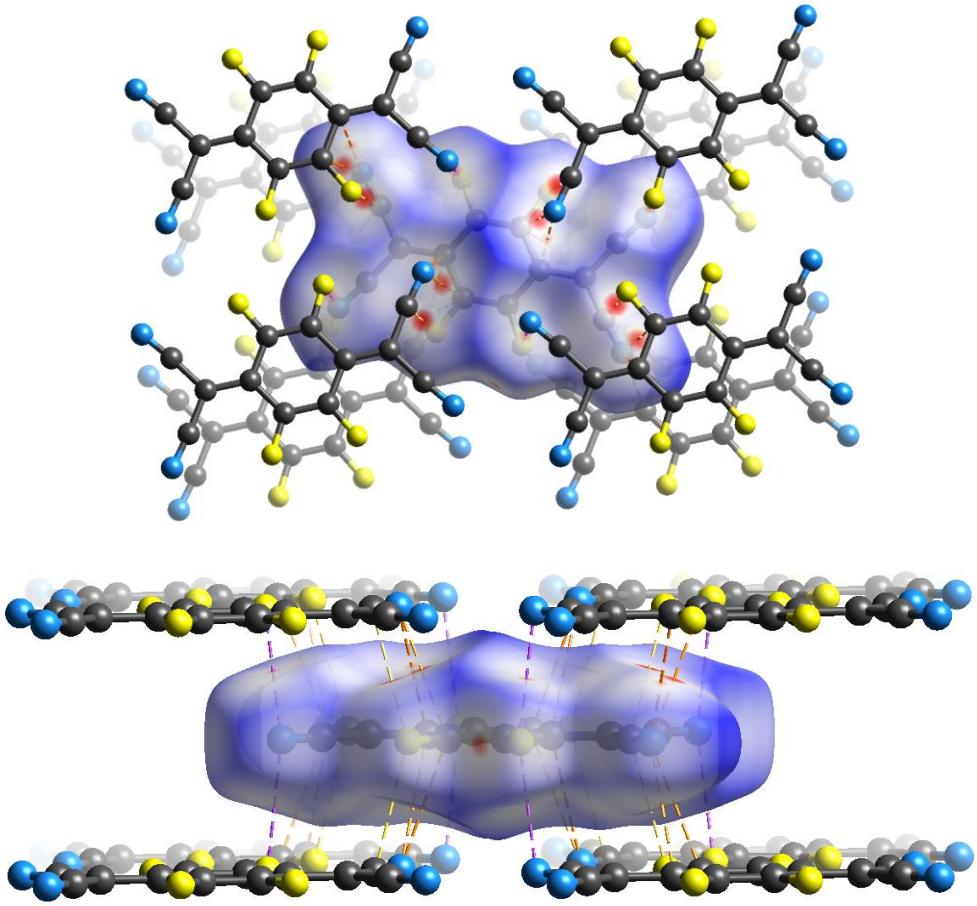
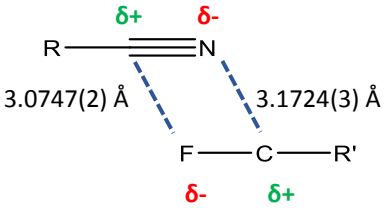


Figure 4.13 Hirshfeld surfaces for II, with adjacent molecules shown. Close contacts are indicated by dashed lines, the colour denoting the atoms which are close; F···N (purple), C···F (yellow), C···N (orange).

One particular pair of contacts creates a 4 membered ring between the CN group of one molecule and the C – F bond of another molecule. Due to the electronegativity difference between atoms within the covalent bonds (e.g. C and N, and F and C) these atoms are likely to be partially charged. The partially charged atoms in the two molecules arrange in such a way that attractive electrostatic interactions could form within a four membered ring, as illustrated in Scheme 4.1. This arrangement is not seen in form I where, although there are

also close contacts between a nitrogen atom and both atoms in a C – F bond of another molecule, the fluorine and cyano carbons are not brought into close contact.



Scheme 4.1 Diagram of C – N···C – F close contact arrangement between molecules in adjacent layers of F_4TCNQ (form II).

Due to this arrangemenent, a neighbouring fluorine within this molecule is then brought into contact with the N of the other molecule, a possibly repulsive interaction, perhaps by virtue of this attractive ring and the stabilisation it provides (Figure 4.14, purple dashed line) in form II.

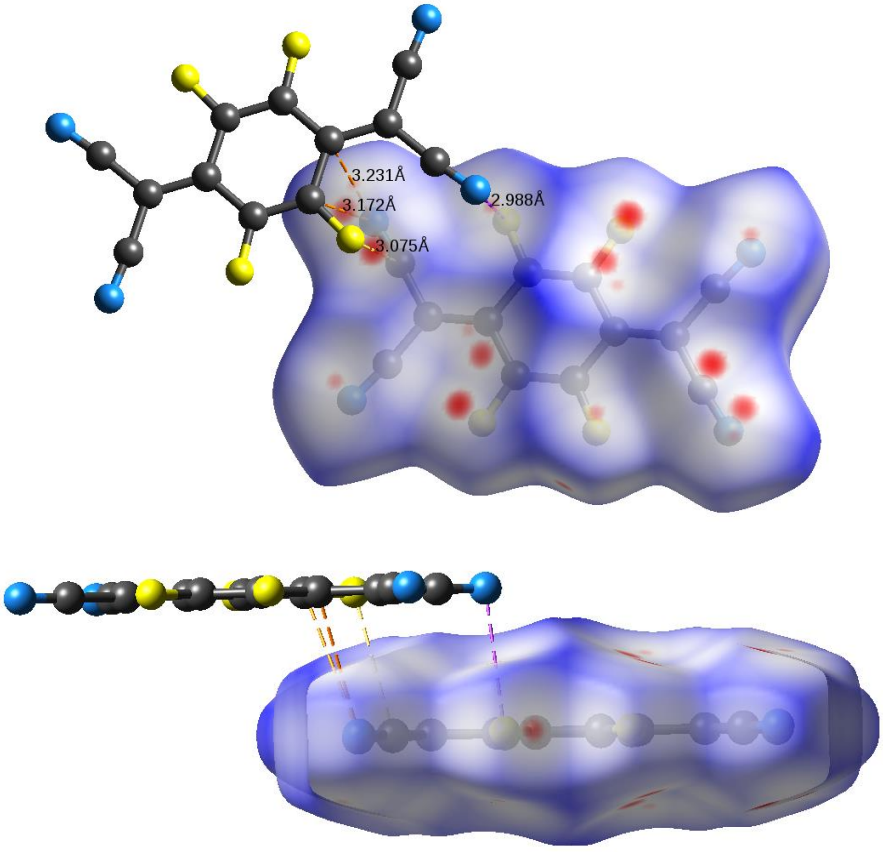


Figure 4.14 Close contact between molecules in adjacent layers illustrated with two molecules.

There is only one type of close contact identified within layers of polymorph II – between fluorine atoms of adjacent molecules (Figure 4.15), an edge-edge interaction. Scientific literature has identified two main types of intermolecular halogen interactions: type I, where the R-X···X angles on each molecule are equal to each other, and type II, where the R-X···X angle ~ 180° (and other ~ 90°). Metrangolo and Resnati state only type II halogen interactions can be classed as a true halogen bond,[154] as in order for a true halogen bond to

occur the nucleophilic portion, or σ hole, of the halogen (along the direction of the R-X bond) must be directed at the electrophilic edge of the other halogen. The R-X···X angle in polymorph II is the same for both molecules (126.92(3)°) which indicates a type I halogen contact. Type I interactions are described as "*van der Waals in nature*".[155] These contacts minimise repulsion between the two halogens[154] and are often deemed to be due to close packing only,[156] although they are suggested to influence crystal packing if both angles are ~ 180°.[157]

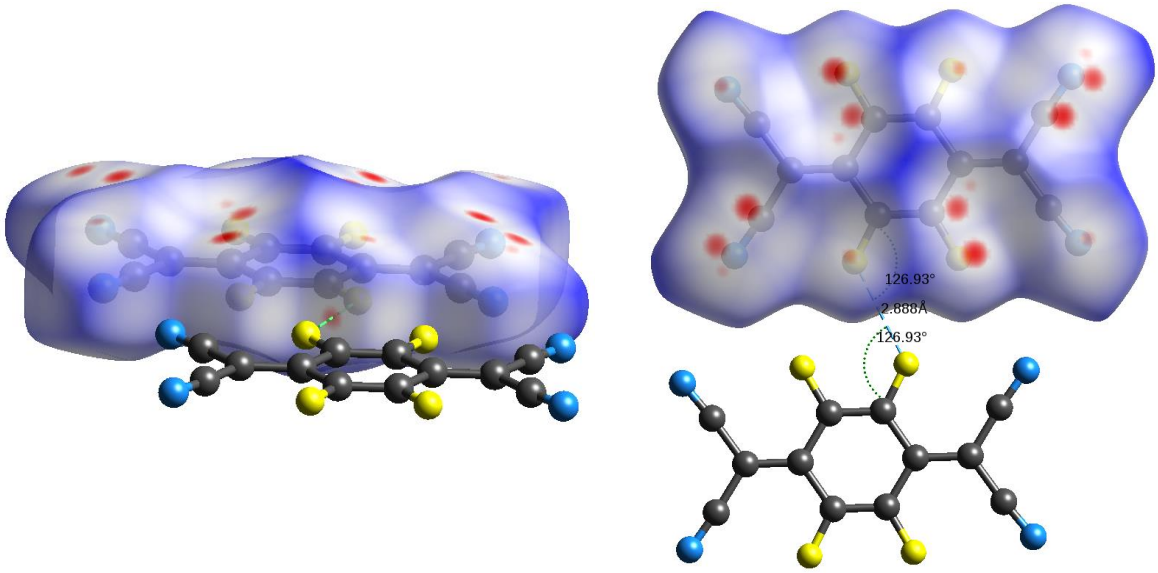


Figure 4.15 Hirshfeld surface for form II showing F···F close contacts within layers.

Fingerprint plots can be calculated to allow further comparison of the Hirshfeld surfaces of the structures.[158] A fingerprint plot provides a graphical summary of the intermolecular distances as displayed on the Hirshfeld surface, illustrating the distance of the closest internal and external nucleus to the Hirshfeld surface (d_i and d_e) at each point on the surface. This allows simple comparisons between crystal structures and has been used to illuminate differences between different crystal polymorphs.[159] Some intermolecular interactions, e.g. hydrogen bonds, produce characteristic features in the fingerprint plot which can be identified. The colours of the plot represent the number of surface points within that particular square (binned at 0.01 Å) on the graph, with a scheme going from blue – green – red indicating an increasing number of surface points at that particular point on the graph.[158]

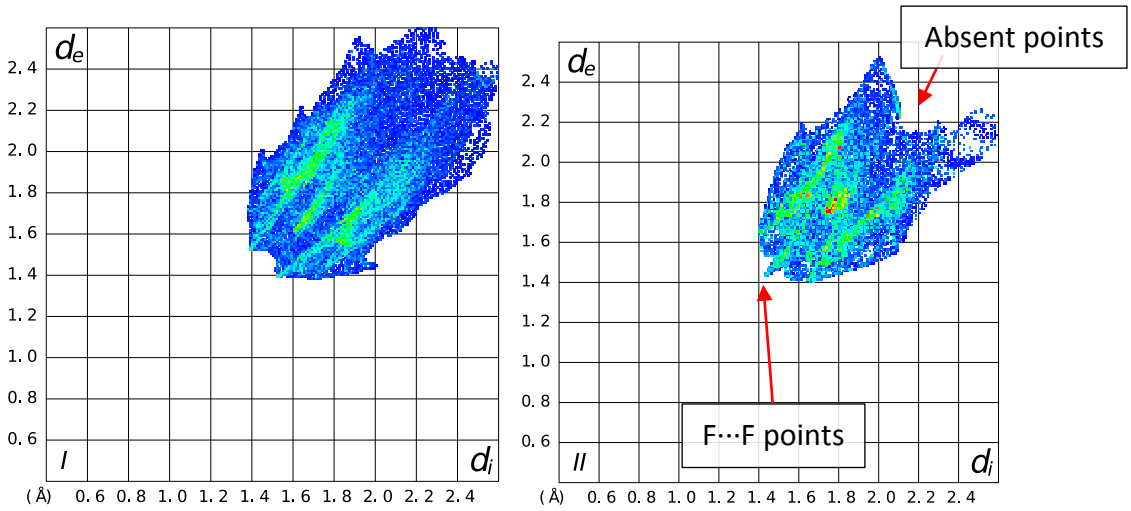


Figure 4.16 Fingerprint plots for polymorphs I and II.

Fingerprint plots for polymorphs I and II (Figure 4.16) illustrate differences between the packing of the two forms. In polymorph I there are more surface points with a longer total distance and a greater spread of points, while in polymorph II there are some additional points along the diagonal of the graph at short distances. Features along the diagonal ($d_i = d_e$) of a fingerprint plot are due to the atom corresponding to d_i and d_e being the equivalent. The point in the plot for form II is a result of the close F···F contacts found within the layers. This is not present in the plot for form I as fluorine atoms in adjacent molecules do not come into as close contact. This is evident when showing only surface points relating to F···F contacts on the fingerprint plot (Figure 4.17).

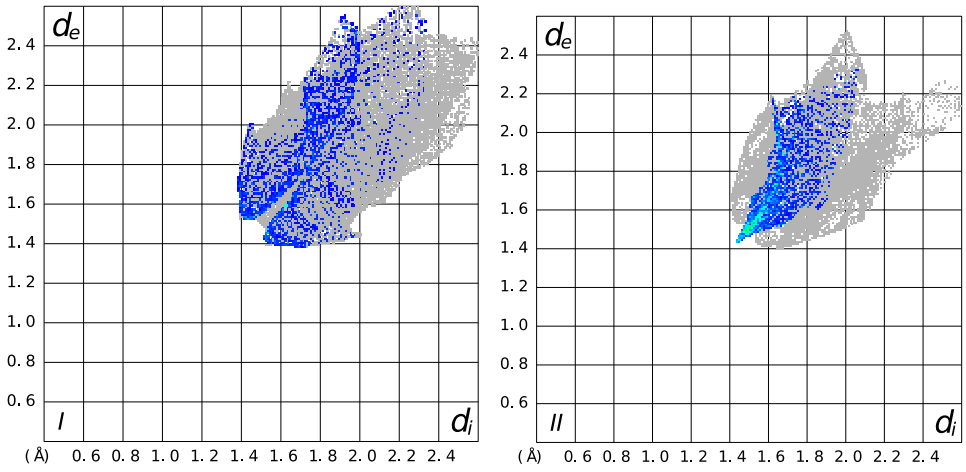


Figure 4.17 Fingerprint plots showing only $F_{internal}\cdots F_{external}$ contacts for polymorphs I and II.

4.5 Energy comparison

As both polymorphs form concomitantly in DCM, but only form I has been reported in the literature, there may be an energetic preference for one form over another. *CrystalExplorer* can calculate pairwise interaction energies between molecules.[160] These can be used to

compute an approximate average energy of the unit cell, and thus determine if one form is more stable than another.

To calculate the average energy of each polymorph, pairwise interaction energies were calculated for molecules present in a radius of 3.8 Å to a central molecule in each case. The calculation of the total interaction energy E_{tot} (Equation 4.1) is made up of the contribution from electrostatics, polarization, dispersion and repulsion each with an associated scale factor to make the values more similar to quantum mechanical calculations, as described by Turner *et al.*[160]

$$E_{tot} = k_{ele}E_{ele} + k_{pol}E_{pol} + k_{dis}E_{dis} + k_{rep}E_{rep} \quad 4.1$$

The scale factors applied for B3LYP/6-3 1 G(d,p) level of modelling are:[161]

$$E_{tot} = 1.057E_{ele} + 0.740E_{pol} + 0.871E_{dis} + 0.618E_{rep} \quad 4.2$$

These total energies for the pairs were then used to calculate the average energy for the crystal structure, as the sum of total energy for N_j equivalent molecules:

$$E_{avg} = \frac{\sum_j N_j \times E_{tot,j}}{\sum_j N_j} \quad 4.3$$

Table 4.3 shows the contributions to the total energy for each independent pairwise interaction energy for polymorph I. The coloured number in column 1 corresponds to the colour of the atoms shown in Figure 4.18, which illustrates the different molecular environments surrounding the central molecule.

N	Symmetry operation	R	E_{ele}	E_{pol}	E_{dis}	E_{rep}	E_{tot}
4	x+1/2, -y+1/2, -z	6.10	-8.3	-4.8	-28.8	24.1	-22.5
2	x, y, z	8.04	-20.5	-3.8	-23.7	17.9	-34.0
4	-x, y+1/2, -z+1/2	8.31	-7.1	-2.0	-15.4	12.1	-15.0
4	-x+1/2, -y, z+1/2	8.60	-3.3	-1.0	-5.6	1.8	-8.0

Table 4.3 Calculated energies between molecules in F_4TCNQ for polymorph I (kJ/mol). Colour of number, N, corresponds to the colour of a molecule within the preceding figure.

There are 4 different molecular environments surrounding the central molecule, two within the layers and two in adjacent layers. Molecules within adjacent layers that are not in the herringbone pattern relative to the central molecule (purple) have the lowest E_{tot}. These molecules also do not have any close contacts to the central molecule. The largest E_{tot} is calculated for molecules within the same layer that are arranged parallel to each other. These molecules exhibit the close C···C contact resulting from overlap of the C(5) atoms with

each other, also shown in Figure 4.10. These molecules have a large electrostatic energy between them – although, as the close contact is between the same atom in both molecules, it is unlikely to be electrostatically favourable. Other factors must contribute to the low electrostatic energy.

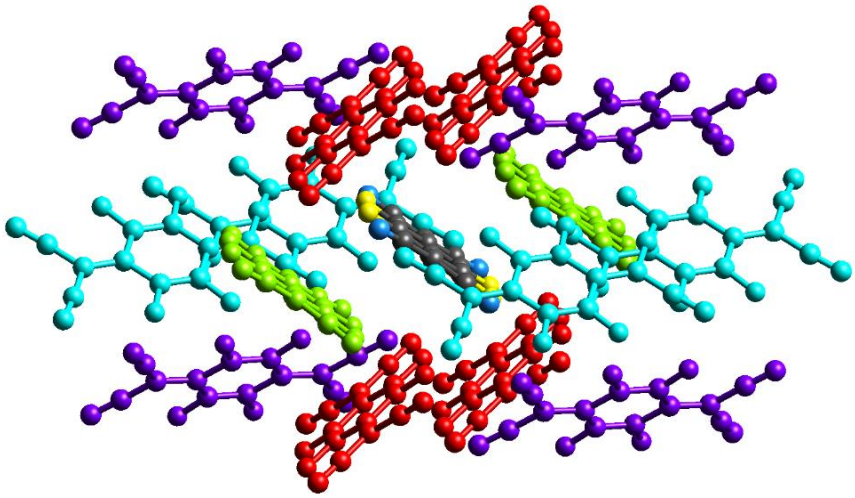


Figure 4.18 Colour coordinated contributions to interaction energy from polymorph I. Colour of molecule corresponds to N in Table 4.3.

Table 4.4 shows the pairwise interaction energies for polymorph II, with the corresponding molecular diagram in Figure 4.19. There are 3 different crystal packing environments around the central molecule, 2 sets within a layer (green and blue) as well as one set of 8 atoms in adjacent layers (red). Molecules in adjacent layers, where in the Hirshfeld surface there were a number of close contacts (Figure 4.13) – including the possibility of a four membered ring of electrostatically attractive atom close contacts (Scheme 4.1), are calculated as having large total energies. Molecules within layers, however, only provide a very small contribution to total energy. Interestingly, the molecules within the layers are the only ones that have a repulsive electrostatic contribution to the total energy in either polymorph. The lowest contribution is from the molecules rendered in green, where in Hirshfeld surfaces there is close contact between the F atoms of adjacent molecules, although the total energy is still slightly attractive (-1.1 $kJmol^{-1}$) due to the contribution from the dispersion energies.

N	Symmetry operation	R	E_{ele}	E_{pol}	E_{dis}	E_{rep}	E_{tot}
2	x, y, z	7.51	8.6	-0.9	-14.1	4.5	-1.1
8	x+1/2, -y+1/2, -z+1/2	7.55	-18.5	-4.1	-26.5	19.9	-33.3
2	x, y, z	11.68	0.5	-0.6	-6.7	3.8	-3.5

Table 4.4 Interaction energies calculated for polymorph II (kJ/mol). Colour of number corresponds to the colour of a molecule within the preceding figure.

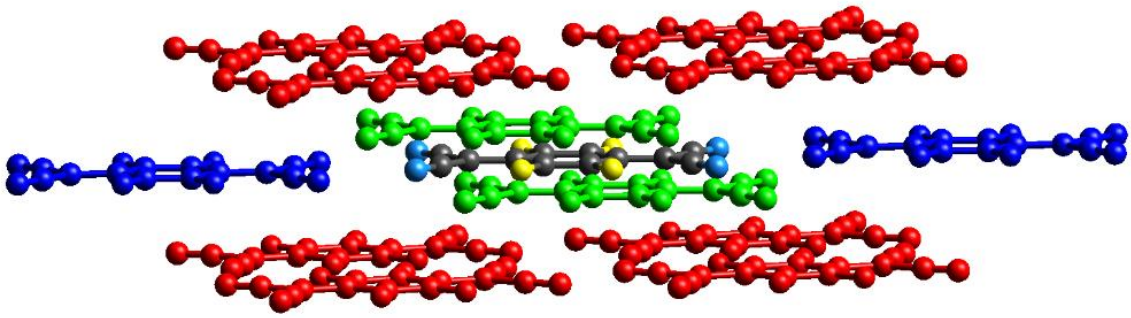


Figure 4.19 Colour coordinated contributions to interaction energy for polymorph II.

The average pairwise energies for form I and II are calculated as -17.85 and -23.03 kJmol^{-1} respectively. These values are similar in energy, which is expected in concomitant polymorphism. The unreported polymorph II is the lower energy thermodynamic polymorph, which raises the question as to why it has not been reported previously.

Crystallisation conditions have been shown to play a role in polymorph formation within scientific literature.[162–164] In the previous reported structures of F_4TCNQ form I that were deposited in the Cambridge Structural Database (CSD),[165] a database of small molecule organic and metal-organic crystal structures, crystals were grown using vapour transport,[166] solution growth,[167] and from a solution of acetonitrile.[168] The appearance of exclusively polymorph I from acetonitrile could suggest an interaction between the solvent and the molecule which prevents or makes it less favourable to form the polymorph II. This may take place in the form of an interaction between the cyano group within acetonitrile with the C – F bond of F_4TCNQ. If acetonitrile blocks other molecules of F_4TCNQ from associating with the C – F bond to form the stabilising 4 membered rings by interacting in that position itself, then other interactions may take precedent during crystallisation.

4.6 Multipolar refinement

The multipole refinement of form II was performed in the same way as the default refinement of form I, as detailed in Chapter 3, using the same order of parameters included in each iteration of the refinement. As the atoms in form II reside on special positions in the unit cell some refinement parameters are symmetry forbidden. The positional and ADP parameters, which are non-vanishing can be obtained using the International Tables,[169] and non-vanishing multipolar parameters are inferred by considering the symmetry of the multipole with respect to the z axis (above and below the molecule). Only multipolar parameters that were non-vanishing with respect to the z axis were refined.

The excluded parameters were thus:

- z
- U_{ij}^{13}, U_{ij}^{23}
- D(0)
- Q(1+), Q(1-)
- O(0), O(2+), O(2-)
- H(1+), H(1-), H(3+), H(3-)

The multipolar refinement strategy from the reference processing of form I from the previous chapter was modified to restrict the refinement of these parameters (Section 10.1.3). After convergence of the refinement strategy, the optimal values for the *a* and *b* parameters for the *SHELXL* weighting scheme were calculated using *CAPOW*. The weighting parameter values were optimised as described in Chapter 3 until convergence of the values occurred to 4 significant figures. The optimal parameters for the weighting scheme were calculated as *a*= 0.0107 and *b* = 0.0062.

4.6.1 Verification of multipolar models

	Polymorph I		Polymorph II	
Weighting	Statistical	*SHELXL*	Statistical	*SHELXL*
# Reflections	4029	4029	2185	2185
R_2 (%)	1.55	1.56	1.29	1.40
R_1 (%)	1.48	1.46	1.20	1.19
wR_2 (%)	1.71	2.75	1.51	2.46
wGooF	1.9668	0.9419	1.9168	1.2281
$d^f(0)$	2.4853	2.4863	2.5028	2.4986
e_{net} (e)	-0.0320	-0.366	0.1200	0.1009
e_{gross} (e)	1.6545	1.6006	1.3100	1.3196
e_{hole} ($eÅ^{-3}$)	-0.174	-0.160	-0.114	-0.113
e_{peak} ($eÅ^{-3}$)	0.204	0.198	0.113	0.117
Data to parameter ratio	11.6783	11.6783	10.1628	10.1628

Table 4.5 Refinement statistics for the statistically and SHELXL weighted multipole modelled form I and form II.

The multipolar models for polymorph II will be compared to the reference and *SHELXL* weighted models from Chapter 3. Global refinement statistics for the two forms are reported in Table 4.5. The R factors for form II are lower than form II, although the wGooF for the statistically weighted models are both comparable and very large. The range of residuals is smaller in form II which could indicate less systematic error within the data. Unlike form I, the application of a *SHELXL* weighting scheme does not result in an overall improvement in refinement statistics as the range of residuals increases, there is a larger

increase in R factor values between the statistical and SHELXL weighted models, and the wGooF is further from 1.

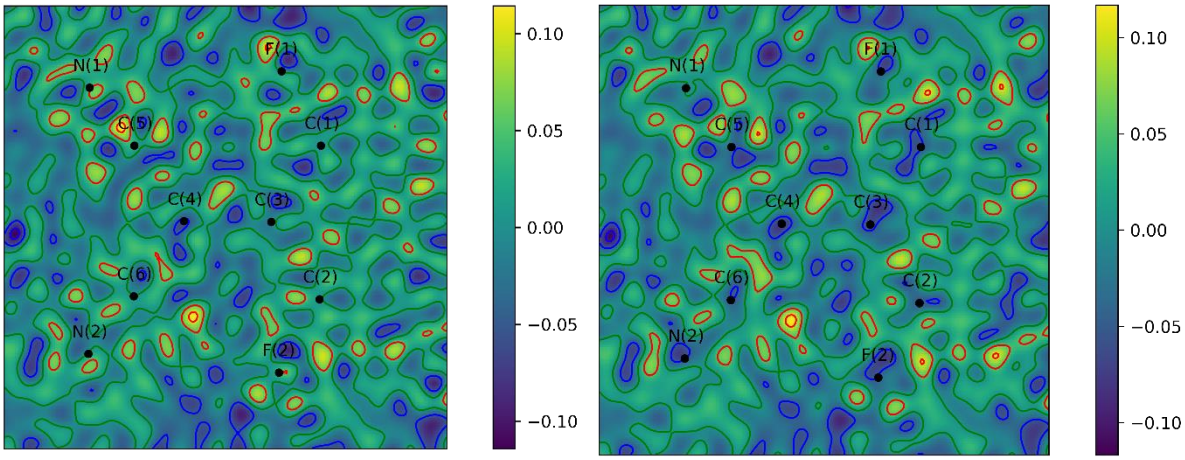


Figure 4.20 2D residual electron density plot for the statistical and SHELXL weighted multipole model of form II of F_4TCNQ calculated in the plane of the molecule. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The zero contour is present in green.

The 2D residual electron density plot in the plane of the molecule (Figure 4.20) is comparable to that of form I (Figure 3.6) with a small amount of residual electron density present. The fractal dimensionality plot for a 3D grid (7 x 7 x 3.5 Å^3) of residual electron density calculated surrounding the asymmetric unit for both the statistically and *SHELXL* weighted multipole model of polymorph II (Figure 4.21) shows a parabolic distribution without features which is an improvement upon that of polymorph I, which is also suggested by the smaller values of e_{net} and e_{gross} for polymorph II.

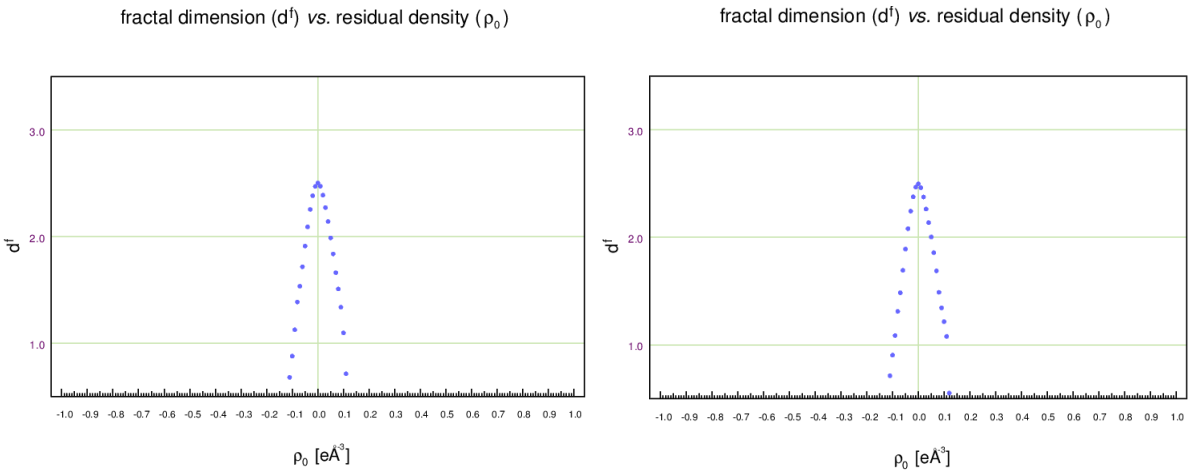


Figure 4.21 Fractal dimensionality plot for statistical (left) and SHELXL (right) weighted multipole model of F_4TCNQ form II.

Again, as in polymorph I, the normal probability plot for the statistically weighted multipole model of polymorph II shows a non-normal distribution of structure factor residuals (Figure 4.22). The application of the *SHELXL* weighting scheme results in a more normal distribution of structure factor residuals and reduces the range of residuals to ± 4. In this case,

improvement of the normal probability plot is not as marked as when the *SHELXL* weighting scheme is applied in form I (Section 3.4.4), shown by residuals outside the range of ± 2 having a greater deviance from the expected normal distribution (red dashed line).

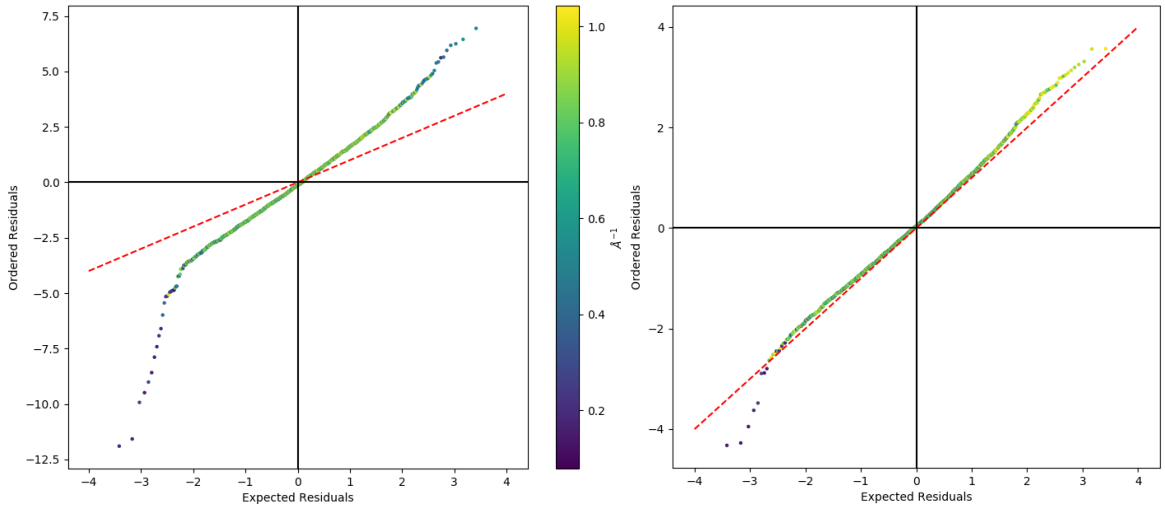


Figure 4.22 Normal probability plot for statistically (left) and SHELXL (right) weighted multipole model of F_4TCNQ form II.

4.6.2 Comparison of multipolar properties

In the case of form II, the application of optimised weights does not result in an improvement of the multipolar model (although there is no statistical difference between multipolar parameters calculated for the statistical and *SHELXL* weighted multipole model of form II). As the weighting scheme does not appear to improve the model of electron density, multipolar properties for form I and II will be compared using the statistically weighted models.

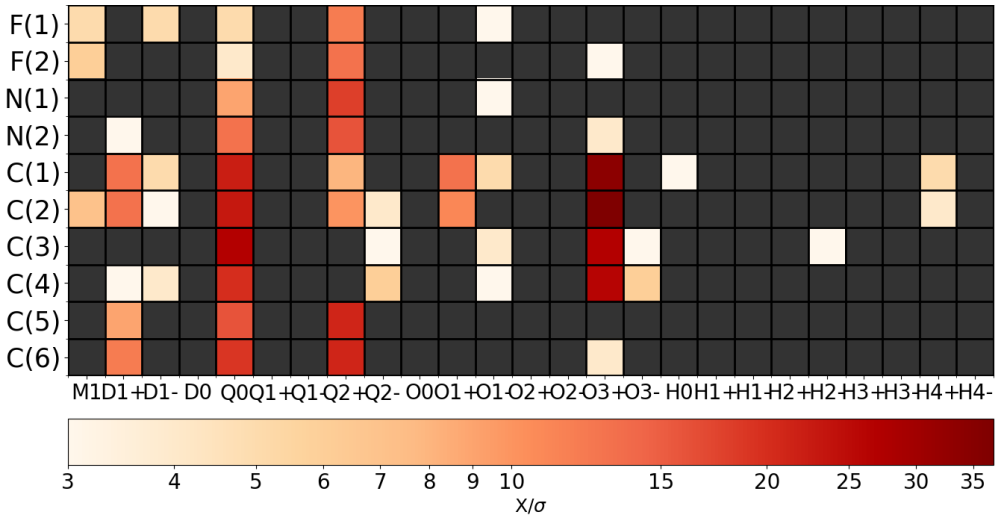


Figure 4.23 Statistical significance of multipolar parameters in statistically weighted model of F_4TCNQ polymorph II. The grid squares in grey show multipole parameters with no statistically significant value.

Figure 4.23 shows the statistical significance of the multipolar parameters within the completed model of polymorph II. This shows that multipolar parameters up to the hexadecapolar level are statistically significant.

The multipolar parameters as reported by *XDLSM* are not directly comparable to those of polymorph I, due to the occupancy difference of the atoms within the asymmetric unit. All atoms in polymorph II have an occupancy of 0.5 (and thus parameters are reported with half their expected value) due to the symmetry constraints of polymorph II. The values of the multipolar parameters must therefore be scaled in order to make the desired comparison.

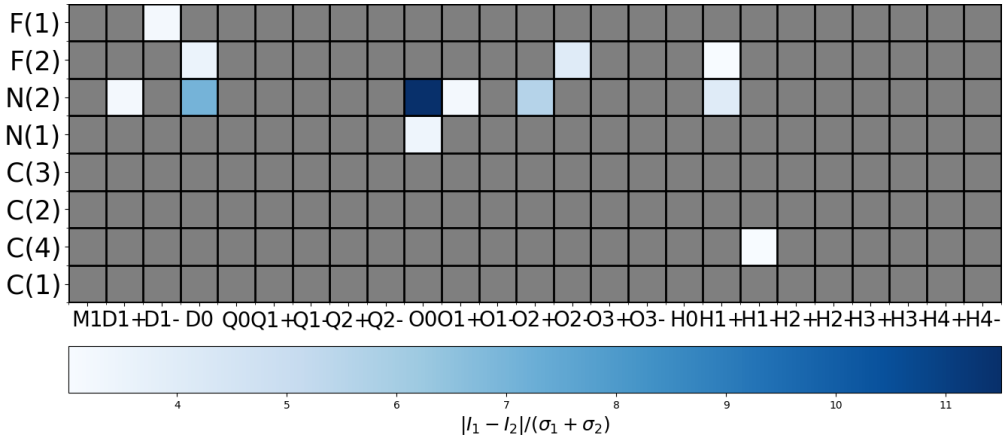


Figure 4.24 Statistically different parameters in polymorph I and II. Squares in grey indicate no statistical difference between parameters in the two molecules.

For most multipolar parameters, there is no statistical difference between the values within the two models (Figure 4.24). Those that do show statistical differences fall into two categories – parameters which are symmetry restricted in form II but have statistically significant values within form I, and those which are non-vanishing in both forms and are statistically different. There are three parameters for which the latter is true, N(2) D(1+), F(1) D(1-) and N(2) O(1+). The value of S_d for these parameters only slightly larger than 3 S_d, therefore these values are only just statistically different. The larger differences in S_d arise from multipolar parameters that were not refined in form II. Overall this shows good general transferability across most of the multipolar models.

		Bond length (Å)	
Atom 1	**Atom 2**	**Form I**	**Form II**
F(1)	C(1)	1.3271(3)	1.3299(3)
F(2)	C(2)	1.3261(3)	1.3299(3)
N(1)	C(5)	1.1577(3)	1.1574(3)
N(2)	C(6)	1.1593(4)	1.1571(3)
C(1)	C(3)	1.4400(3)	1.4396(3)
C(2)	C(3)	1.4398(3)	1.4412(3)
C(1)	C(2*)	1.3515(3)	1.3516(3)
C(3)	C(4)	1.3768(3)	1.3776(3)
C(4)	C(5)	1.4287(3)	1.4285(4)
C(4)	C(6)	1.4296(3)	1.4268(3)

Table 4.6 Bond lengths of form I and form II.

A comparison of the intramolecular bond lengths within the two structures shows differences between the two models (Table 4.6). The F(1) – C(1), F(2) – C(2), N(2) – C(6), and C(4) – C(6) bonds are statistically different, although this may be caused by the high degree of precision that the bond lengths are reported to and thus the very low standard uncertainties, as opposed to the values being genuinely different.

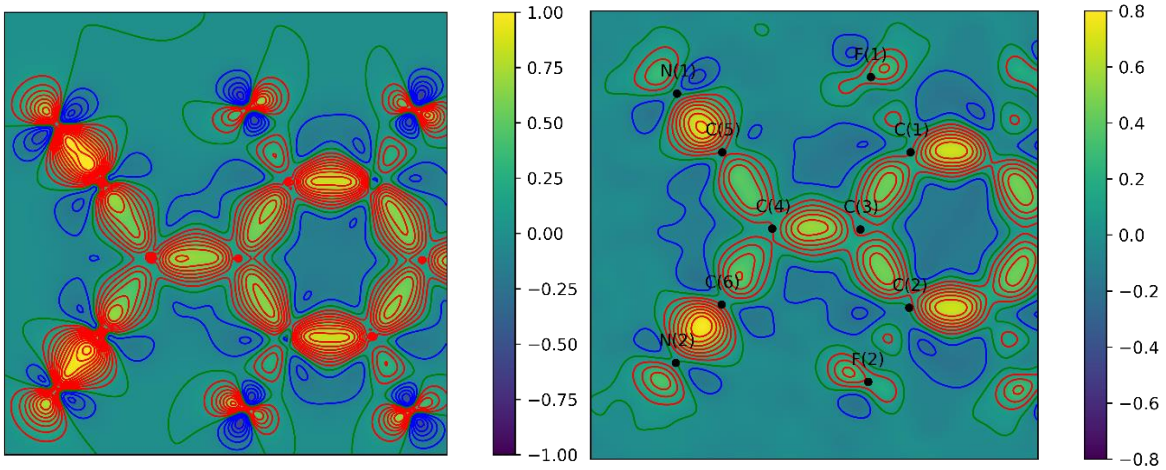


Figure 4.25 Static and dynamic deformation electron density for form II of F_4TCNQ. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The zero contour is present in green. The colourbar for the static deformation density plot is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

The static deformation electron density calculated in the plane of the molecule for form I (Figure 3.10 and Figure 3.11) and form II (Figure 4.25) have a broadly similar distribution of density, which would be expected if the distribution of electron density within the molecule was the same in each polymorph. The dynamic deformation density plots are also similar.

		Form I			Form II		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.135(15)	-23.381(85)	0.1	2.100(8)	-21.761(45)	0.11
F(2)	C(2)	2.096(15)	-21.754(86)	0.08	2.078(8)	-22.653(47)	0.13
N(1)	C(5)	3.397(25)	-28.932(199)	0.06	3.431(12)	-35.467(98)	0.01
N(2)	C(6)	3.373(24)	-29.683(201)	0.06	3.406(13)	-32.304(109)	0
C(1)	C(3)	2.011(8)	-16.645(30)	0.18	1.987(5)	-16.593(18)	0.18
C(2)	C(3)	2.009(9)	-17.119(30)	0.16	2.004(5)	-17.642(18)	0.17
C(1)	C(2*)	2.391(26)	-23.629(67)	0.39	2.416(15)	-23.993(38)	0.41
C(3)	C(4)	2.209(10)	-20.075(35)	0.27	2.212(6)	-19.152(21)	0.26
C(4)	C(5)	1.955(9)	-14.829(31)	0.11	1.946(6)	-14.383(19)	0.09
C(4)	C(6)	1.913(9)	-13.004(32)	0.08	1.892(5)	-13.627(19)	0.10

Table 4.7 Comparison of intramolecular BCPs between form I and form II of F_4TCNQ.

The properties at BCPs for form II are in good agreement with those of form I (Table 3.6) and display a similar trend in the relative values of ρ and $\nabla^2\rho$, with the highest values and largest differences between the two forms recorded at the BCP for the CN triple bond. The values of standard uncertainty for ρ and $\nabla^2\rho$ are all lower for form II over form I, which could be due in part to the higher significance of the reflections measured for form II. There are

systematic differences between values of the Laplacian in the two polymorphs, showing a greater charge accumulation within the cyano group bonds (but no larger than that observed for differing processing within Chapter 3).

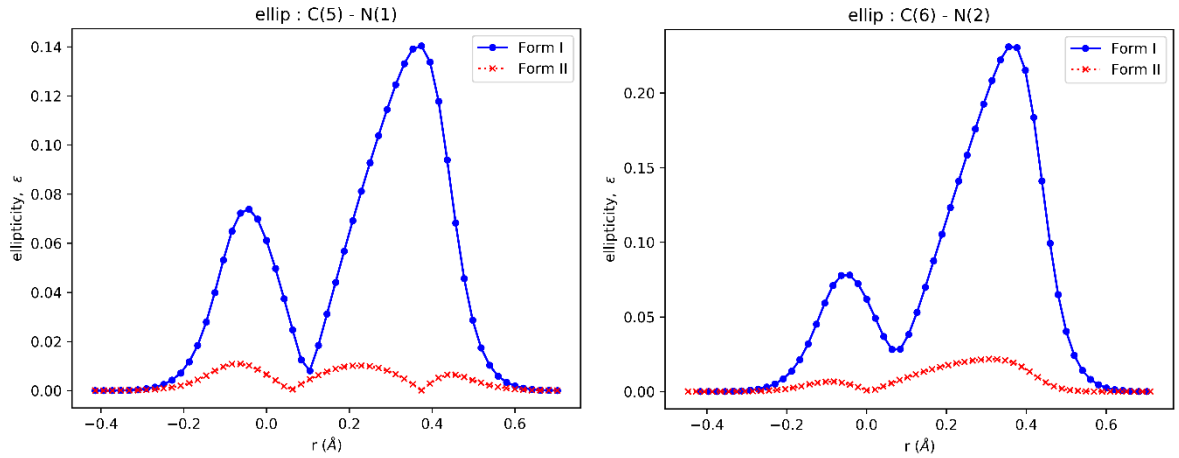


Figure 4.26 Ellipticity along bond path for cyano group bonds in form I and II of F_4TCNQ.

There are differences in the ellipticities, which also transfer to the ellipticity when calculated along the bond path. While the overall shape of the ellipticity remained similar for most bonds, there were large differences in that of the C(5) – N(1) and C(6) – N(2) bonds (Figure 4.26). These plots show broadly the same features in the two forms but the value of ε are considerably different. This may be due to the inclusion of multipoles which are symmetry forbidden in form II; as discussed in Section 3.4.4, ε along the bond path is a very sensitive property and the variation along the bond path observed may be within the expected standard uncertainty of this property. The differences did not extend to the ρ and $\nabla^2\rho$ plots, which had the same function in both forms.

	Form I				Form II			
Atom	P_{val}	κ	κ′	Net atomic charge	P_{val}	κ	κ′	Net atomic charge
F(1)	7.101(11)	0.999	0.953	**-0.100(11)**	7.068(12)	1.000(1)	0.959	**-0.067(12)**
F(2)	7.097(11)	0.999	0.953	**-0.096(11)**	7.078(12)	1.000(1)	0.959	**-0.078(12)**
N(1)	5.108(25)	0.996(2)	1.009	**-0.108(25)**	5.024(28)	0.999(2)	1.001	-0.024(28)
N(2)	5.058(26)	0.996(2)	1.009	-0.057(26)	5.051(28)	0.999(2)	1.001	-0.051(28)
C(1)	3.908(19)	1.016(2)	0.985	**+0.092(19)**	3.946(21)	1.014(2)	0.986	+0.054(21)
C(2)	3.883(19)	1.016(2)	0.985	**+0.117(19)**	3.843(21)	1.014(2)	0.986	**+0.157(21)**
C(3)	3.982(21)	1.010(2)	0.979	+0.017(21)	3.964(23)	1.010(2)	0.982	+0.035(23)
C(4)	4.084(22)	1.010(2)	0.979	**-0.083(22)**	4.059(24)	1.010(2)	0.982	-0.059(24)
C(5)	3.856(28)	1.039(2)	0.990	**+0.143(28)**	3.964(31)	1.031(3)	0.992	+0.036(31)
C(6)	3.924(29)	1.039(2)	0.990	**+0.075(29)**	4.003(29)	1.031(3)	0.992	-0.002(29)

Table 4.8 Net atomic charges for form I and form II.

While there is no statistical difference between the net atomic charges or P_{val}, the differences in absolute value should still be noted. There are more statistically significant values (emboldened within Table 4.8) in form I than form II. These values have an expected

sign for the charge. The value for P_{val} is > 4 for C(6) which results in a positive net atomic charge, the opposite of what would be expected due to the negative charge of N(2) – however, this is not a statistically significant value.

4.7 Pseudo-polymorphism of F_4TCNQ

Slow evaporation of a saturated solution of F_4TCNQ in toluene produces red needle-shaped single crystals. Data for this crystal was measured using Cu wavelength (λ = 0. 1.54184 Å) at 150 K on a Gemini Atlas CCD detector in order to provide an initial measurement to identify if the crystal was a new polymorph of F_4TCNQ. The unit cell was found to be different from the previous two polymorphs and therefore the data were processed to obtain a spherical atom structural solution in order to assess for the suitability of charge density refinement.

The frames were integrated and scaled using *CrysAlisPro* and structure solution and refinement took place using *XL* and *XT* respectively. Structure solution and refinement of the collected data yielded a structure containing both F_4TCNQ and toluene molecules, also known as a solvate – as solvent molecules are incorporated into the crystal structure. The structure can be described as a pseudo-polymorph of F_4TCNQ; unit cell parameters are listed in Table 4.9. The crystal packs in the centrosymmetric *$P2_1/c$* space group.

Space group	*$P2_1/c$*
Lattice type	Monoclinic
a (Å)	8.1314(3)
b (Å)	7.4141(2)
c (Å)	13.6796(4)
β (°)	100.551(3)
V (Å^3)	810.76(4)
ρ (g/cm^3)	1.509
Z	2
Z’	0.5

Table 4.9 Unit cell parameters for F_4TCNQ-toluene solvate.

The structure is comprised of discrete homomolecular layers, parallel to the crystallographic [1 0 0] plane, with the rings of the toluene and F_4TCNQ molecules stacked in an alternating sequence (Figure 4.27). Within the crystal structure, the toluene molecules are modelled as being disordered over two positions. Because of this disorder, the crystal was not considered suitable for charge density refinement. The arrangement of the two molecules suggests π- π interactions between the rings of the two molecules. Due to the inversion of the dipole within the ring of F_4TCNQ (as F is more electron withdrawing than C, whereas H is less) as opposed to toluene, this could be a favourable interaction.

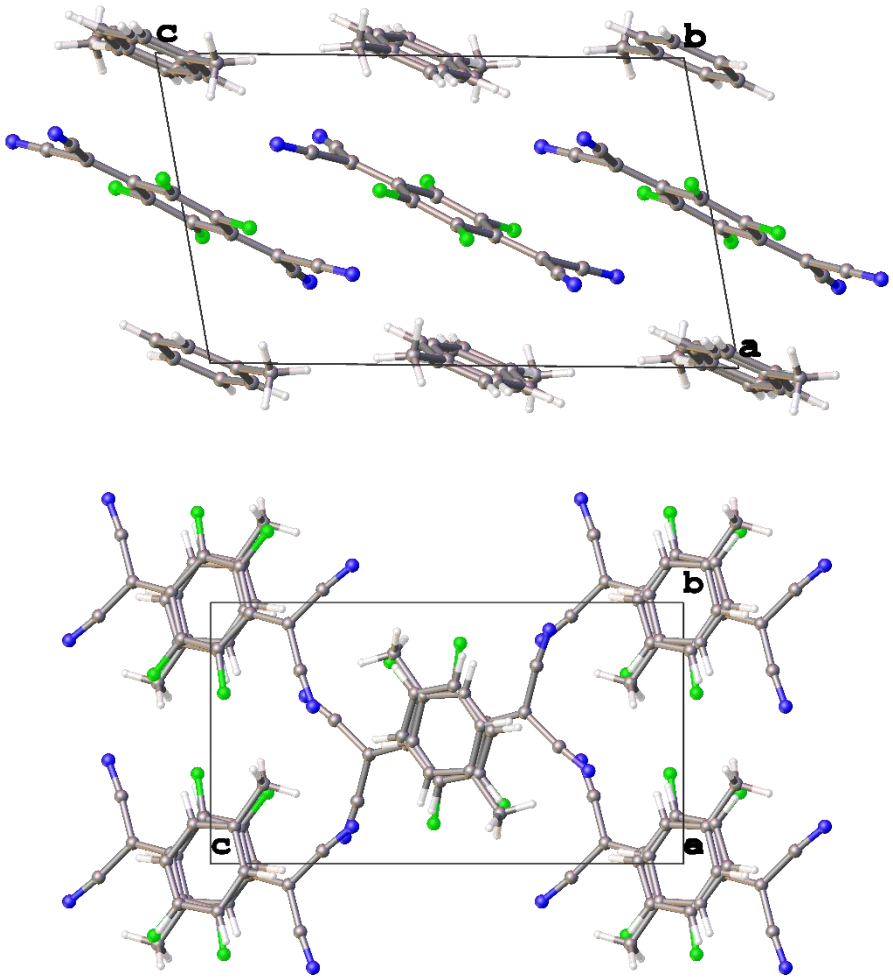


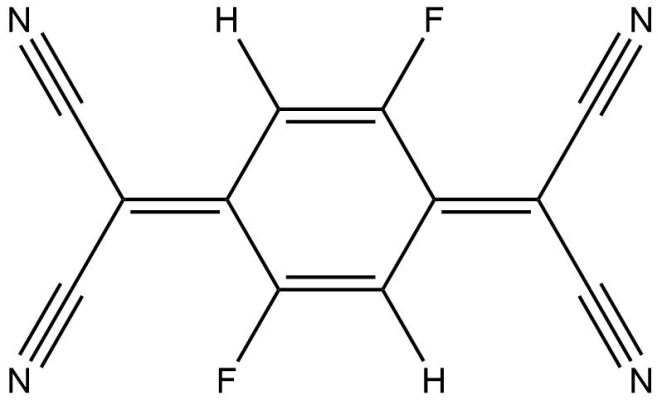
Figure 4.27 Crystal packing of F_4TCNQ-toluene solvate.

4.8 Structural comparison of polymorph II to F_2TCNQ

F_2TCNQ has a layered structure, similar to form II of F_4TCNQ (Figure 4.28). The layered structure of F_2TCNQ is said to give rise to the crystals electron transport properties.[170–172] The family of F_nTCNQ (n=0,2,4) structures has been studied for the potential use in organic semiconductors;[166] F_2TCNQ (2,5-difluoro-7,7,8,8-tetracyanoquinodimethane, $C_{12}H_2F_2N_4$, Scheme 4.2) was found to have a much higher electron mobility (6-7$cm^2V^{-1}s^{-1}$ at room temperature and as large as 25$cm^2V^{-1}s^{-1}$ at 150K) compared to values of 0.1 and 0.2$cm^2V^{-1}s^{-1}$ for TCNQ (n=0) and F_4TCNQ (n=4) respectively, when measured at room temperature. Unlike TCNQ and F_4TCNQ, F_2TCNQ also exhibits "*band-like*" electron transport in a field effect transistor – which is shown by a reduction in electron mobility when the temperature is lowered.

Chernyshov *et al*. suggest that having a small number of molecules within the primitive unit cell (Z_{red} = 1) leads to a number of forbidden low-frequency optical vibration modes.[170] The electron-phonon interaction is weakened by the lack of these modes and therefore an

increased electron mobility results. The investigation of form II is required to assess whether it possesses the same properties, however, comparison of the two structures is important with a view to this end.



Scheme 4.2 2,5-difluoro-7,7,8,8-tetracyanoquinodimethane (F_2TCNQ).

The reported IAM structure of F_2TCNQ was analysed in a similar way to the structures of F_4TCNQ within this chapter. The crystal structure and atom positions used were taken from Krupskaya *et al.*[166] as the data were measured at 100 K. F_2TCNQ crystalises in the centrosymmetric *C2/m* spacegroup. Unit cell parameters are reported in Table 4.10.

Space group	*C2/m*
Lattice	Monoclinic
a (Å)	10.1746(8)
b (Å)	5.8549(4)
c (Å)	8.8403(6)
β (°)	106.939(8)
V (Å)	503.78(7)
ρ (g/cm^3)	1.583
Z	2
Z’	0.25

Table 4.10 Unit cell parameters for F_2TCNQ.

The main difference between F_2TCNQ and form II is that molecules in adjacent layers are not twisted with respect to each other as seen when comparing Figure 4.29 to Figure 4.8; this prevents the four membered ring CN···CF interaction as seen in form II. Instead weak hydrogen bonds of the type CH···NC are formed within layers. The hydrogen bonds may override the formation of the CN···CF ring and thereby direct the formation of the crystal structure.

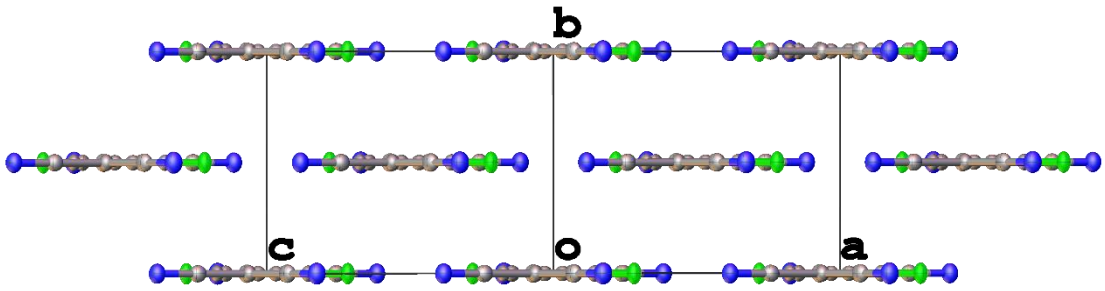


Figure 4.28 A view of F_2TCNQ along the [1 0 1] axis showing the formation of layers of molecules.

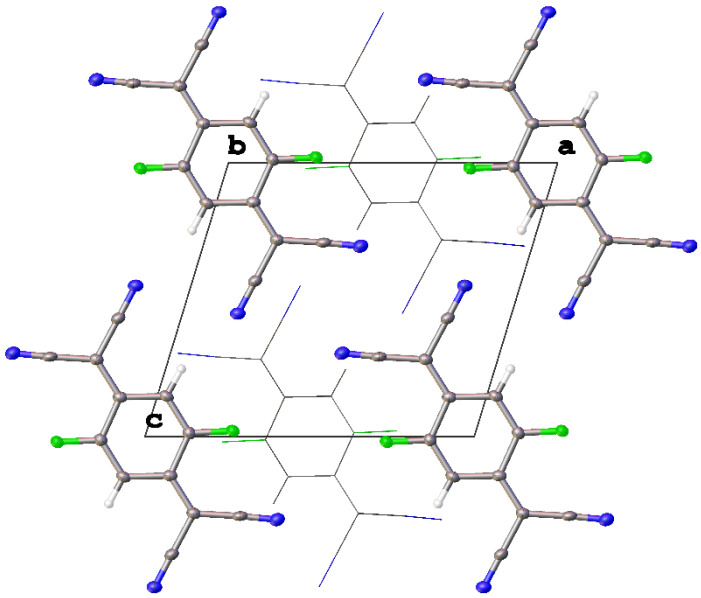


Figure 4.29 View of F_2TCNQ along [0 1 0] axis. Molecules in adjacent layer are drawn with wireframe model.

Hirshfeld surfaces created for F_2TCNQ are shown in Figure 4.30. Figure 4.31 illustrates that the only close contacts within the layers are hydrogen bonds involving the cyano nitrogen.

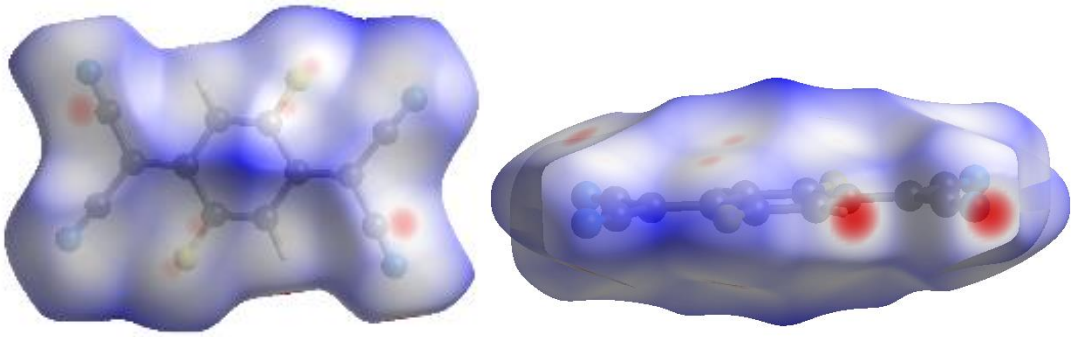


Figure 4.30 View of Hirshfeld surfaces of F_2TCNQ, showing d_{norm} values where red areas indicate where two atoms are closer than the sum of their van der Waals radii.

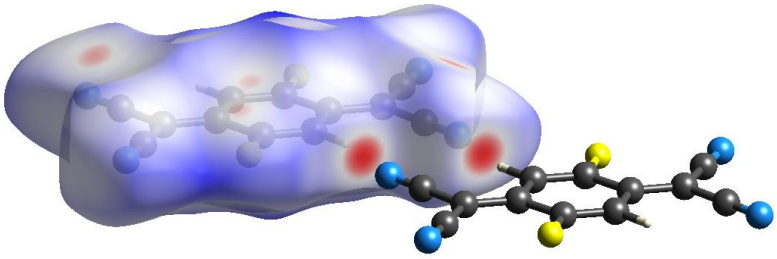


Figure 4.31 Hirshfeld surface showing hydrogen bonding within layer.

Between layers in F_2TCNQ, although there are no 4 membered CN···CF close contacts, there is a similar close contact ring between atoms within the C – F bonds of adjacent molecules (Figure 4.32). There is a similar C···C close contact observed in the cyano carbons, with additional close contacts identified between C and N, potentially forming another stabilising 4 membered ring.

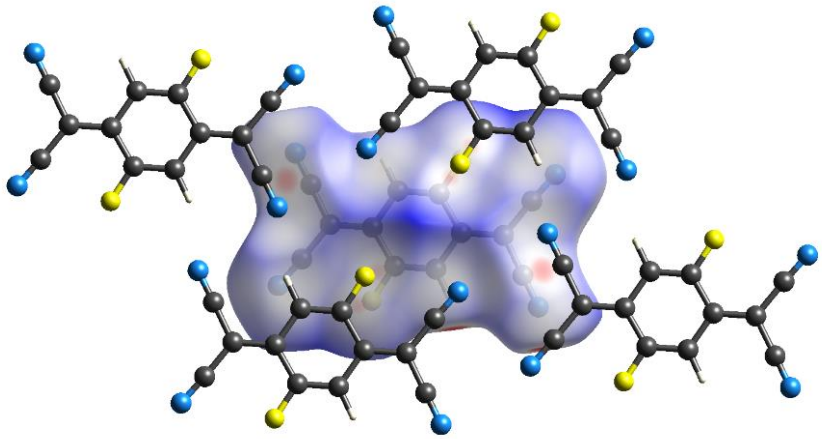


Figure 4.32 Close contacts between layers of F_2TCNQ.

The fingerprint plot calculated for F_2TCNQ (Figure 4.33) shows a different distribution of intermolecular contacts. The characteristic two points for hydrogen bonding are present within the plot (Figure 4.34). There is some evidence of F···F close contacts at a similar distance in F_2TCNQ (Figure 4.35), but these are not as numerous as they are in form II.

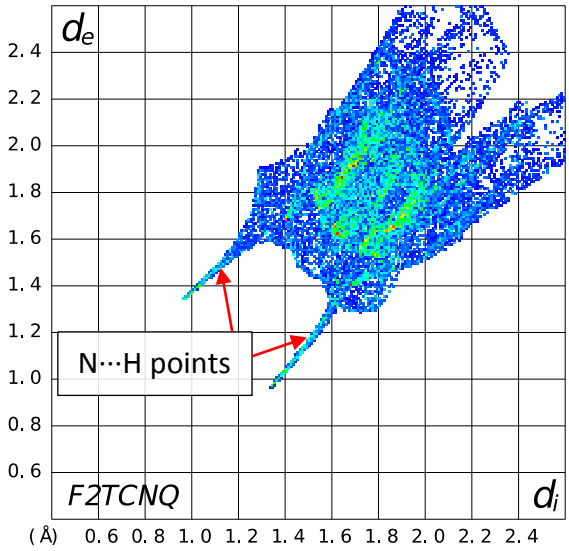


Figure 4.33 Fingerprint plot of F_2TCNQ.

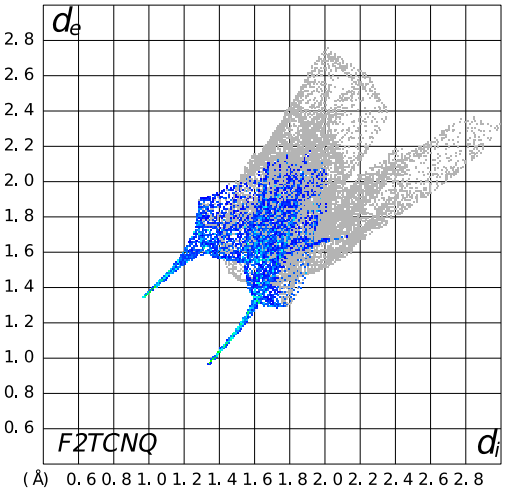


Figure 4.34 Fingerprint plot for F_2TCNQ showing only N···H contacts.

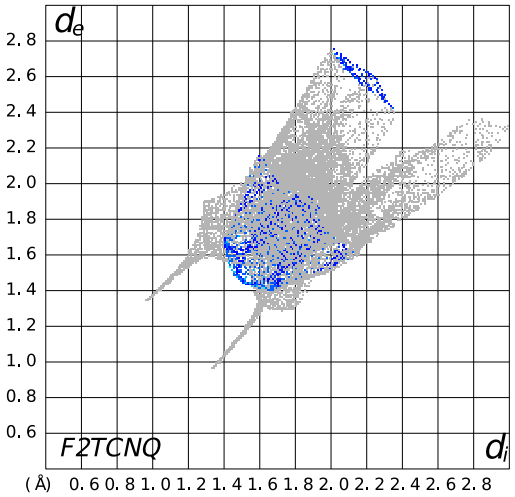


Figure 4.35 Fingerprint plot for F_2TCNQ showing only F···F contacts.

When pairwise interaction energies are calculated for the structure in *CrystalExplorer*, the green molecules, illustrated in Figure 4.36, have the largest total energy − 29.0 kJmol^{-1} (Table 4.11). This is comparable to − 33.3 kJmol^{-1} as calculated for molecules in adjacent layers of form II. However, there is a much greater contribution to the energy of F_2TCNQ from molecules within the layers, owing to the hydrogen bonding present within the structure.

N	Symmetry operation	R	E_{ele}	E_{pol}	E_{dis}	E_{rep}	E_{tot}
2	x, y, z	11.37	-1.8	-1.0	-8.9	7.8	-5.6
4	x+1/2, y+1/2, z	5.87	-9.2	-4.9	-33.2	21.6	-29.0
2	x, y, z	8.84	-28.4	-4.5	-12.1	28.2	-26.4
4	x+1/2, y+1/2, z	9.29	-11.8	-1.9	-16.3	16.3	-18.0

Table 4.11 Pairwise interaction energies for F_2TCNQ (kJ/mol). Colour of N in table corresponds to colour of a molecule within the following figure.

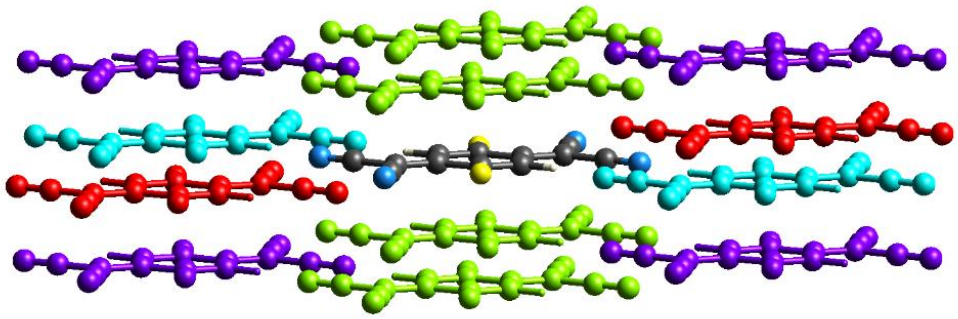


Figure 4.36 Colour coded contributions to pairwise interaction energy for F_2TCNQ.

4.9 Conclusions and future work

These results show the evidence of polymorphism in F_4TCNQ. This is of importance as a number of published pieces of work comment on the fact no evidence of polymorphism of the molecule was found.[166] Both forms of F_4TCNQ are very stable under ambient conditions. Over a period of 6 months no interconversion between forms was observed.

Different polymorphs of structures can have different properties, despite the molecular structure remaining the same.[173] The new polymorph, II, of F_4TCNQ exhibits a layered structure which has been suggested to promote electron mobility and charge transfer.[170] Raman spectroscopy performed with FnTCNQ (n=0,2,4) crystals showed a correlation between the lowest observed vibrational frequency in the Raman spectrum and the charge mobility of the single crystal; the lowest vibrational mode for F_2TCNQ was almost double the values detected for TCNQ and F_4TCNQ (polymorph I). The Raman spectrum of polymorph II could therefore illuminate whether polymorph II also exhibits similar properties to F_2TCNQ.

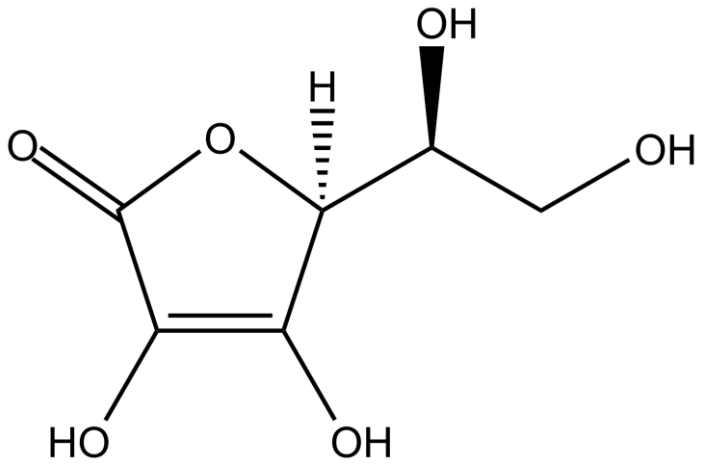
The growth of a TCNQ/F_4TCNQ co-crystal was attempted. This was postulated due to the co-crystallisation of both hexadeuterobenzene with hexafluorobenzene and F_4TCNQ with toluene show a similar alternating layer structure. However, there was also no evidence found of a co-crystal forming between TCNQ and F_4TCNQ (in saturated solutions of either acetonitrile or DCM). There was also no evidence found of an equivalent toluene solvate with TCNQ. All crystals that were tested that were produced from toluene solutions were of TCNQ alone.

Overall, there was little statistical difference between the multipolar parameters for form I or II where both parameters are non-vanishing in form II, which suggests that multipolar parameters are transferable between the two molecules. This is encouraging as the principle of transferability of multipolar parameters underlies the TAAM. Further studies could be performed to probe the multipole models for the two polymorphs more thoroughly. Comparisons of intermolecular bonding through the calculation of BCPs (alongside the study of the IAM using Hirshfeld surfaces) could result in a deeper understanding of the formation of the two polymorphs.

Chapter 5. The transferability of multipole parameters in L-ascorbic acid

In Chapter 4, the polymorphism of F_4TCNQ was exploited in order to probe the transferability of multipole parameters between atoms in crystallographically similar environments. Another class of crystal structures that would allow the internal testing of transferability are homomolecular crystals with more than one molecule in the asymmetric unit (i.e. where Z' > 1). Crystal structures with Z' > 1 exist throughout the crystallographic literature and their occurrence has been studies extensively.[147,174–179] There appears to be no singular reason for a crystal structure to have Z' > 1; this feature has been attributed to crystallisation conditions,[147] crystal packing effects[177] or the structure of the molecule and its conformational flexibility.[179]

Having multiple molecules in the unit cell allows comparison of multipolar parameters for atoms in similar environments using the same experimental dataset, which removes data processing or data collection as an independent source of error between the molecules. Therefore, any variation between atoms in different molecules should be attributable to a difference in the crystallographic environment of the atoms, unless systematic errors are present.



Scheme 5.1 L-ascorbic acid (vitamin C).

One such compound is L-ascorbic acid (vitamin C, $C_6H_8O_6$, Scheme 5.1), in which Z' = 2. The charge density of this compound has previously been studied in 1997 by Milanesio *et al.*[20]

No values for the multipolar parameters were reported for any atoms in the published work, therefore, transferability cannot be investigated with the results currently available. The availability of good crystals of L-ascorbic acid make this molecule an attractive candidate for further investigation. Further to this, advances in equipment and computational techniques mean that it may be useful to reanalyse previously studied materials in the hope of gaining new insights into their structure. From a biological perspective, L-ascorbic acid is incredibly important to human life. A deficiency of L-ascorbic acid can lead to many health complications, including scurvy which can ultimately cause death.[180]

5.1 Data processing

L-ascorbic acid (99%) was purchased from Sigma Aldrich and used without further purification. Single crystals were produced by slow cooling a saturated solution of distilled water and L-ascorbic acid over a period of 24 hours. A crystal of size 0.23 x 0.24 x 0.32 mm^3 was selected and transferred to a *MiTeGen* loop and mounted upon the goniometer. This crystal was slowly cooled to 100 K at a rate of 1 K per minute. Data were collected by with a Bruker D8 venture diffractometer with a Photon II CPAD detector, using an Incotec Iμs 3.0 Ag source (λ = 0.56086 Å) which included a 100 μm aluminium filter, at the Bruker Applications science laboratories by Dr H. Ott in collaboration with Newcastle University. The data collection strategy used by Bruker AXS GmbH was one optimised in order to collect the highest possible quality data for charge density refinement in 72 hours (Section 10.3.1).

The diffraction frames were integrated in *SAINT* using the default integration algorithm with a best-plane background and an experimental cut-off of 0.45 Å. The data were scaled and merged in *SADABS* using the default error model, a correction for overloaded reflections and a numerical correction based on the faces of the crystal. The space group was identified in *XPREP* and the IAM solution and refinement were performed in the *OLEX2* GUI using *XT* and *XL* respectively. An extinction correction was applied to the data to correct for secondary extinction (due to multiple scattering of diffracted X-ray beams within a crystal weakening the overall diffracted intensity of reflections).[146] Throughout all refinements, atoms are labelled as indicated in Figure 5.1.

L-ascorbic acid crystallises in the non-centrosymmetric monoclinic $P2_1$ space group. The unit cell parameters (Table 5.1) calculated for L-ascorbic acid are statistically different from those reported by Milanesio *et al.*[20] possibly due to the different temperatures the data were measured at (100 K and 120 K). A standard uncertainty for the volume for Milanesio *et al*. is

not provided as it was not present in the published work. There appears to be a contraction in both the *b* and *c* axes while there is an increase in the length of the *a* axis when the temperature is lowered. This could be due to the large number of hydrogen bonds, mostly within the *a* – *c* plane. Hydrogen bonds which are highly directional and the reduction in temperature may result in rearrangement of molecules to accommodate the preferred geometries.

	Thesis	**Milanesio *et al.*[20]**
Temperature (K)	100(2)	120
a (Å)	6.3987(3)	6.390(1)
b (Å)	6.2486(3)	6.262(1)
c (Å)	17.1084(7)	17.127(4)
β (°)	99.318(1)	99.36(2)
V (Å^3)	675.02(5)	676.20

Table 5.1 Unit cell parameters of published and 100K L-ascorbic acid data.

A summary of refinement statistics for the spherical atom model are reported in Table 5.2. The spherical atom modelled data has a low R_1 factor of 2.09%, which is an improvement on that of the Milanesio modelling (3.8%). The residuals reported from the spherical atom model display an appreciable amount of unmodelled electron density located predominantly in between atoms, where electron density for covalent bonding may be expected.

R_1 (%)	2.09
wR_2 (%)	6.22
R_{int} (%)	3.27
GooF	1.101
e hole (eÅ^{-3})	0.546
e peak (eÅ^{-3})	-0.302
Extinction correction	0.010(5)
Weighting scheme	*SHELXL* a = 0.037, b = 0.026

Table 5.2 Summary of refinement statistics of L-ascorbic acid after IAM refinement.

The arrangement of atoms within the two molecules in the asymmetric unit (Figure 5.1) of the spherical atom model is broadly similar, with slight differences in the orientation of the atoms within the ethane diol moiety of the molecule. Figure 5.2 shows a diagram of the two molecules overlaid on each other (produced by matching 3 atoms within the 5 membered ring of each molecule in *Olex2*), which further illustrates this difference.

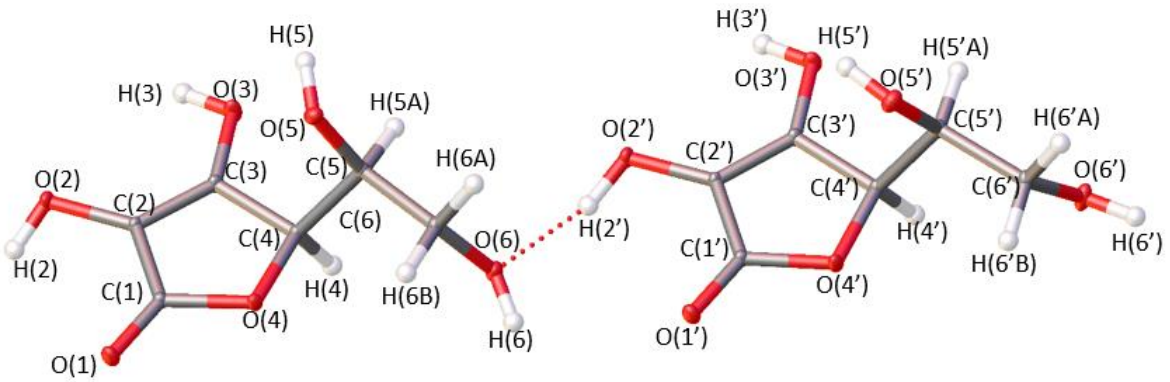


Figure 5.1 Asymmetric unit of l-ascorbic acid. molecule 1 (left), molecule 2 (right) - ' after the atom number indicates the atom is located in molecule 2.

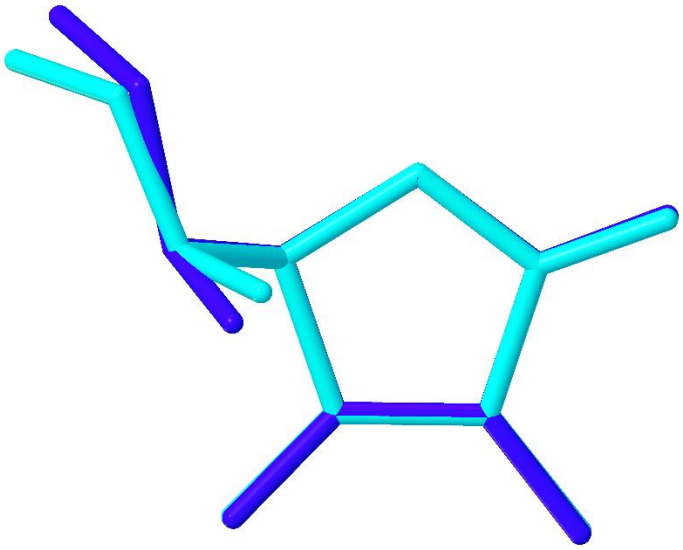


Figure 5.2 Overlay of non-hydrogen atoms in asymmetric unit (molecule 1 = dark blue, molecule 2 = light blue).

In the L-ascorbic acid structure each molecule forms 8 hydrogen bonds, utilising all hydroxyl groups in the molecule. This is consistent with Milanesio *et al.*[20] Figure 5.3 shows these hydrogen bonds involving the atoms in the asymmetric unit, indicated by dashed lines. The atom labelled in italics indicate the atom, of a different molecule to the ones shown, involved in the hydrogen bond. The geometries of these hydrogen bonds differ between the 2 independent molecules, as do the identities of the acceptor oxygen atoms which the hydrogen atoms interact with. This is summarised in Table 5.3.

Hydrogen atom	Molecule 1	Molecule 2
H(2)/H(2')	O(5)	O(6)
H(3)/ H(3')	O(1)	O(1')
H(5)/ H(5')	O(6')	O(2')
H(6)/ H(6')	O(5')	O(2)

Table 5.3 Hydrogen bonding acceptors for molecules 1 and 2 of L-ascorbic acid.

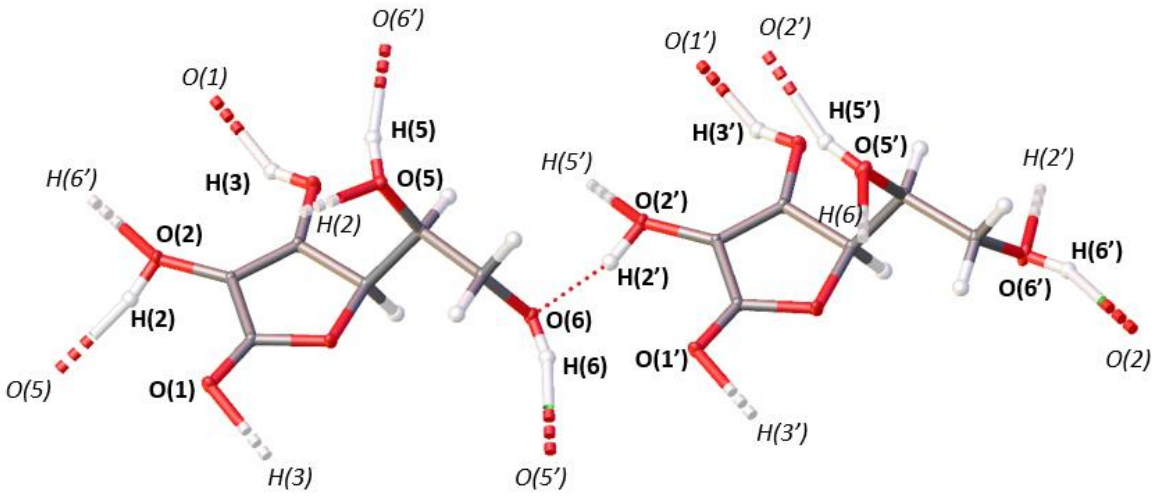


Figure 5.3 Asymmetric unit of L-ascorbic acid showing hydrogen bonding, labels in italic are those from symmetry equivalent molecules.

5.2 Multipolar analysis and transferability of parameters

Multipolar refinement was performed using *XD* with the least-squares minimisation calculated against F^2. The direction of the local coordinate system for each atom was carefully chosen to be approximately the same for equivalent atoms in both molecules in order to facilitate a true comparison of multipolar parameters. Multipoles up to the hexadecapolar level were included for non-hydrogen atoms, while the monopole and one bond directed dipole parameters were refined for hydrogen atoms. Reflections with an $I/\sigma_I < 3$ were excluded from the data processing, as it was shown in Chapter 3 that the application of an I/σ_I cut-off at this level did not result in statistically different values for multipolar parameters. κ and κ' values were refined for chemically similar atoms as summarised in Table 5.4. The full refinement strategy is described in Section 10.3.3.

κ group	Atoms
1	O(2)/O(2')/O(3)/O(3')
2	O(1)/O(1')
3	O(4)/O(4')
4	O(5)/O(5')/O(6)/O(6')
5	C(2)/C(2')/C(3)/C(3')
6	C(4)/C(4')
7	C(5)/C(5')/C(6)/C(6')
8	C(1)/C(1')
9	All hydrogen atoms

Table 5.4 The atoms pertaining to each κ/κ' parameter in L-ascorbic acid multipolar refinement.

A number of reflections were excluded from the refinement due to white radiation contamination. The white radiation contamination was not apparent until the application of a *SHELXL* weighting scheme using *CAPOW*. An initial precession image created using all of

the experimental runs to check for the presence of the effect was clear of the characteristic streaks. The outliers within the normal probability plot of the statistically weighted data, which were used to identify the presence of the contamination for TCNQ (Chapter 2), were within the range of ± 20 which was observed for other statistically weighted data sets within this thesis.

When optimised values for the *a* and *b* parameters of the weighting scheme were calculated and applied, the resulting normal probability plots still showed the presence of large negative residuals instead of the expected reduction of the range of residuals, which was observed in Chapter 3. Reflections identified as the largest outliers displayed a trend in their *hkl* indices (0 *k l* reflections, where *k* = *l*). Upon location of the reflections within the collected diffraction frames, several reflections with a non-Gaussian distribution of intensity were identified (see Chapter 2). The creation of synthesised precession images without the fast scans included finally showed the characteristic streaks within the data. Those reflections that were worst affected, the largest outliers within the normal probability plot, were removed from the *.hkl* file and the multipolar refinement was performed with these reflections omitted. Further information and the precession images created are included in Section 10.3.2.

The treatment of hydrogen atoms is based on the best practise as presented by Hoser *et al*.[126] in their study of the effect of different techniques for the processing of hydrogen atoms on charge density models. Anisotropic atomic displacement parameters were calculated for hydrogen atoms using the *SHADE* (Simple Hydrogen Anisotropic Displacement Estimator) webserver[130] – an online program which estimates the U_{ij} parameters for these atoms. It has been shown that when these parameters are neglected and an isotropic model of atomic displacement is used for hydrogen atoms, that all topological features of a charge density model can be affected, not just those for hydrogen bonds.[129,131] The *SHADE* server calculates a translation-libration-screw model[181] for the rigid non-hydrogen atom framework of the molecule and extends this model to hydrogen atoms along with a contribution from the expected internal motion of the hydrogen atom (derived from a database of neutron studies) in order to provide a best estimate of the hydrogen atom U_{ij} values. Hydrogen atom positions were located using a refinement of the hydrogen position against only low resolution reflections and the bond length was extended and constrained to the expected bond lengths from neutron diffraction studies.[126]

Unlike previous charge density refinements, an extinction correction was required to be included in the refinement strategy or convergence of multipolar parameters did not occur. There are several different types of correction which can be applied in *XD* – the default correction (type 1, Lorentzian) was applied in this case. Without an extinction correction the normal probability plot created from the refinement shows many low angle reflections with large negative residual values.

After convergence of the refinement, optimal values for the *a* and *b* parameters of a *SHELXL* weighting scheme were calculated using *CAPOW*. The weighting parameter values were optimised as described in chapter 3 until convergence of the values occurred to 4 significant figures. The optimal parameters for the weighting scheme were calculated as a = 0.0117 and b = 0.0059.

5.2.1 Verification of multipole model

The refinement statistics from both the statistical and *SHELXL* weighted multipole refinement are listed in Table 5.5. The same trend in refinement statistics described in Chapter 3 is observed with the inclusion of a weighting scheme for the L-ascorbic acid data with an increase in wR2 and a decrease in GooF. In this case, the e_{net} and e_{gross} are both smaller which, along with the decrease of the range of the residual electron density, implies the weighted model has a reduction in the unrefined density and thus an improved model of the electron density. The wGooF was calculated as 1.6923, a value much greater than 1, which is a common trend highlighted in previous chapters. This was reduced to 1.0237 with the inclusion of the *SHELXL* weighting scheme.

	Statistical weight	*SHELXL* weight
Total reflections	15459	15459
Reflections used	15040	15040
R_2 (%)	1.58	1.59
R_1 (%)	1.09	1.07
wR_2 (%)	1.80	2.43
wGooF	1.6759	1.0237
$d^f(0)$	2.5091	2.5096
e_{net} (e)	-0.2065	-0.1839
e_{gross} (e)	8.3217	8.0095
e hole ($eÅ^{-3}$)	-0.102	-0.101
e peak ($eÅ^{-3}$)	0.141	0.134
Extinction correction	0.051(2)	0.046(2)
Nref/Nv	17.5906	17.5906

Table 5.5 Refinement statistics for statistical and SHELXL weighted multipolar refinements of L-ascorbic acid.

Residual electron density plots for both weighting schemes show the same distribution of electron density in the plane of the ring for both molecules and a mostly flat and featureless distribution, as pictured for the *SHELXL* weighted model in Figure 5.4 (also in statistically weighted plots shown in Figure 10.37). This featurelessness is further reflected in the fractal dimensionality plots (calculated using a 3D grid of residual electron density which encloses atoms within the asymmetric unit, created using *XDFOUR*). Plots for the statistical and *SHELXL* weighted model are approximately the same (Figure 5.5).

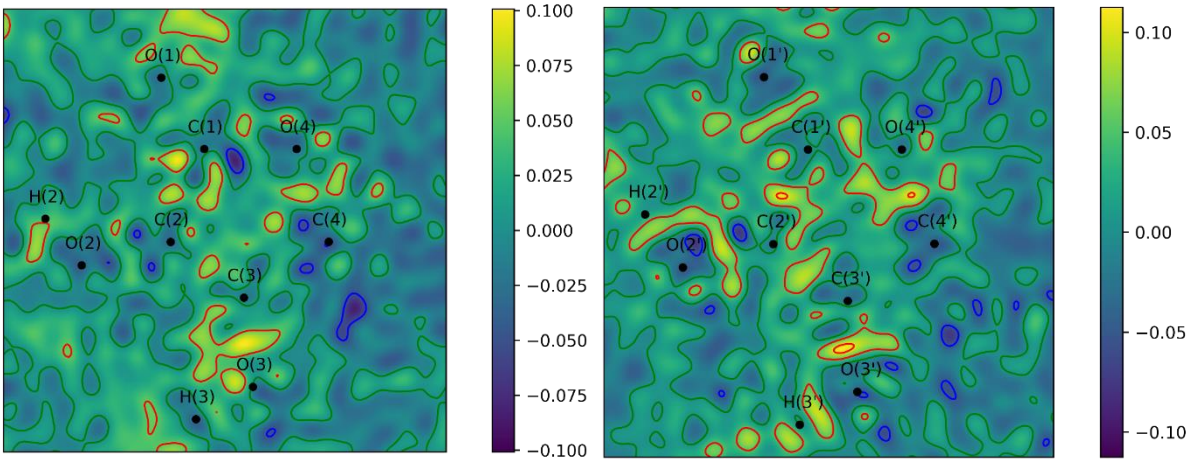


Figure 5.4 Residual electron density plots in plane of ring for molecule 1 and 2 for SHELXL weighted model of L-ascorbic acid. Contours are drawn at ± 0.05 eÅ^{-3} levels, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green.

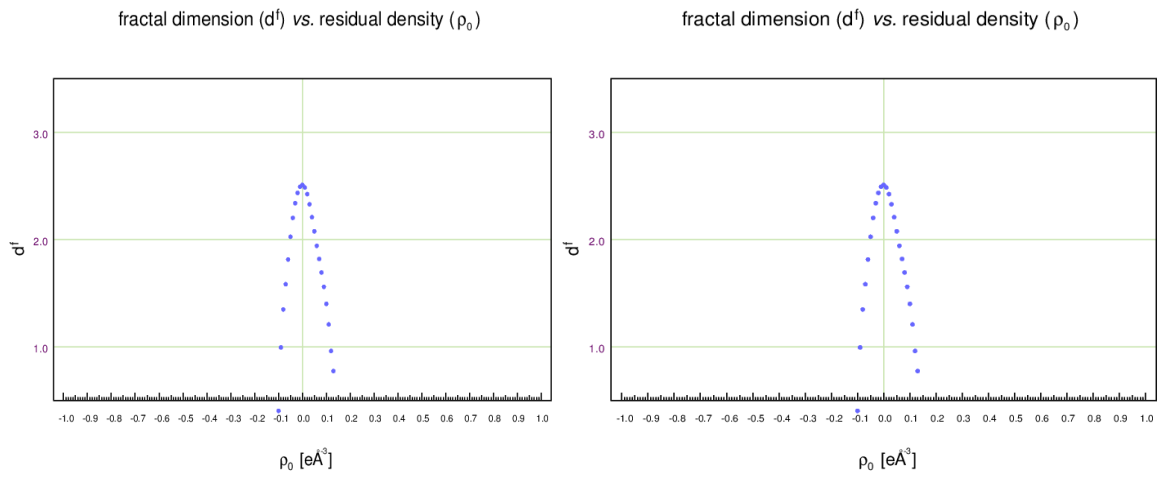


Figure 5.5 Fractal dimensionality plot of residual electron density of statistical (left) and SHELXL (right) weighted multipole model of L-ascorbic acid.

The normal probability plot calculated with *CAPOW* indicates a non-normal distribution of residuals when statistical weights are utilised (Figure 5.6). When a weighting is applied there are still reflections at low extremities which may indicate the extinction correction applied does not fully take into account the extinction effect. Other forms of the extinction correction can be applied but these appear to make little difference to the overall form of the normal probability plot. The application of the *SHELXL* weighting scheme results in a

more normal distribution of residuals with a reduced range of almost entirely between ± 4, as shown in Figure 5.7.

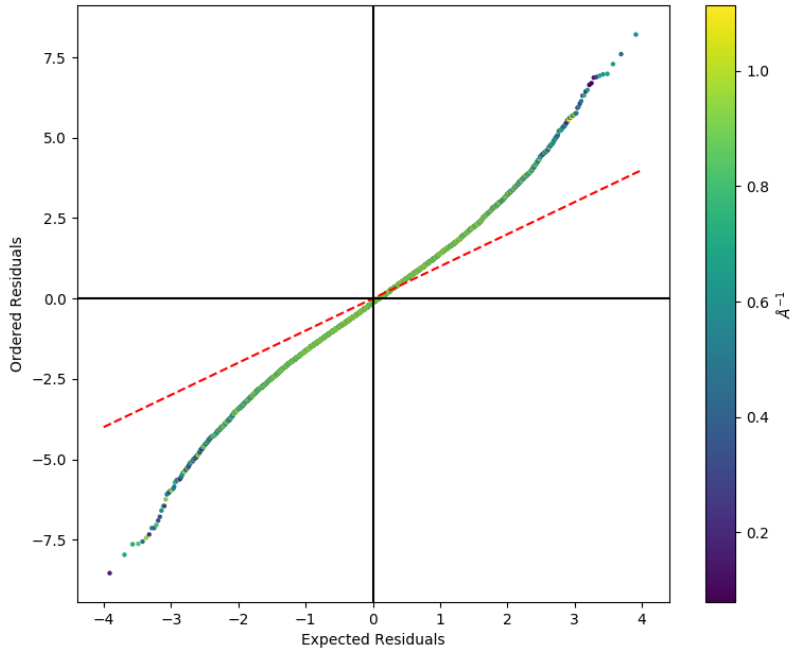


Figure 5.6 Normal probability plot for statistically weighted L-ascorbic acid multipolar refinement.

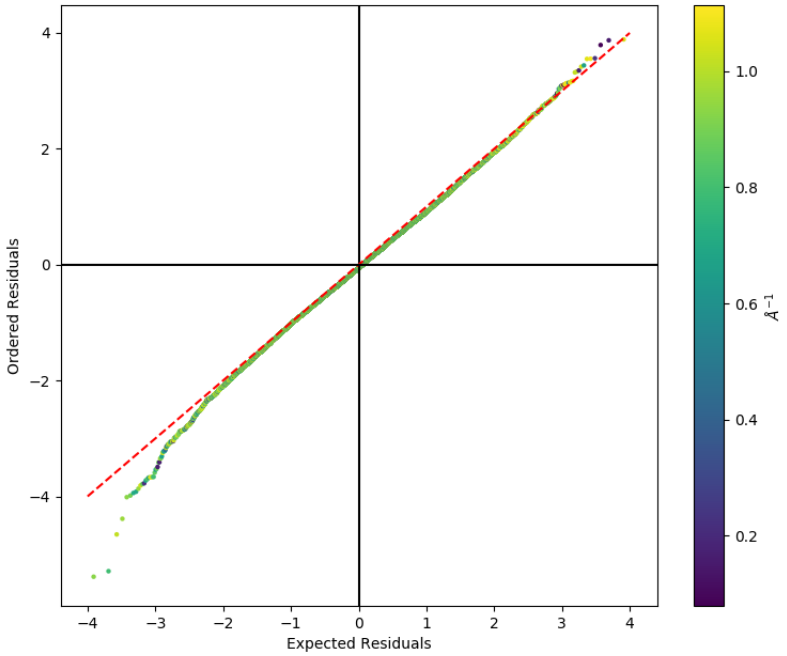


Figure 5.7 Normal probability plot of SHELXL weighted multipole model for L-ascorbic acid.

Additionally, the difference of the mean square displacement in the direction of bonding atoms showed that all non-hydrogen bonds passed the Hirshfeld rigid bond test. The multipole models in both cases are therefore deemed to be reasonable; the transferability of the multipole parameters for the statistical and *SHELXL* weighted models is assessed in the next section.

5.2.2 Transferability of multipolar parameters

The transferability of the multipolar parameters is investigated by calculating whether values for the multipolar parameters of equivalent atoms in both molecules are statistically different from each other. Statistical difference is determined if the ranges of $x \pm 3\sigma$ (where x is the value of the multipole parameter) calculated for the same value of multipolar parameter in both molecules does not intersect. This is analysed graphically using a plot of statistical difference of multipolar parameters (Figure 5.8). The colour scheme corresponds to the value of S_d (Equation 3.1) calculated for the two parameters where $S_d > 3$. All multipolar parameters which do not reach this criterion are shown as grey squares.

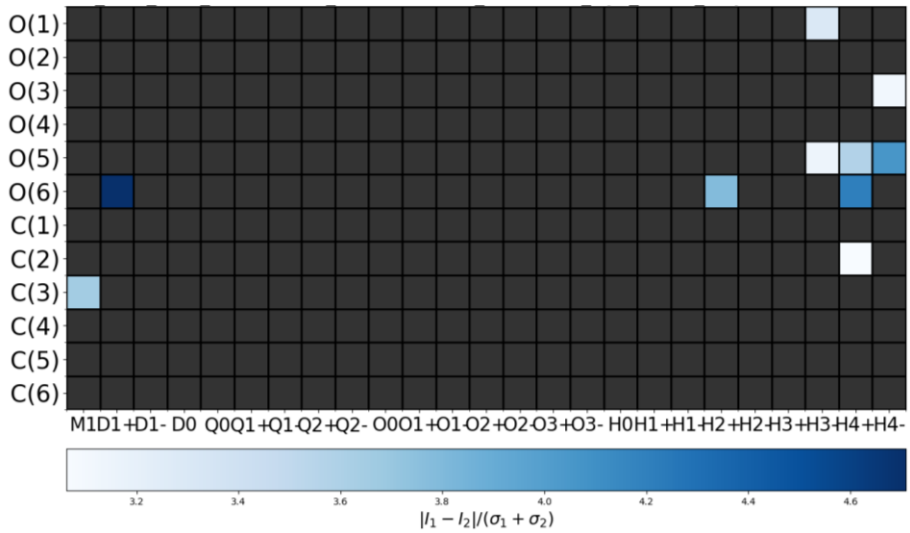


Figure 5.8 Grid indicating multipolar parameters which ranges ± 3σ do not overlap between equivalent atoms in both molecules in L-ascorbic acid.

Most multipole parameters are not statistically different between the two molecules by this definition of statistical difference. The largest differences in S_d are visible for O(5) and O(6). The values are greater than 4 S_d (with a maximum of 4.42 for O(6), D1+) which is larger than an 99.99% confidence interval that the value for the parameters are the same if they are assumed to be normally distributed. These differences in multipolar parameters for O(5) and O(6) atoms may be caused by the difference in geometry of the ethane diol group in the two molecules (Figure 5.2) resulting in the direction of the multipolar axes not being comparable (as they are assigned using vectors to other atoms within the molecule). Alternatively, this could be an experimental feature caused by the different geometry of intermolecular hydrogen bonding interaction for these atoms which, along with different intermolecular bond lengths, could result in a different distribution of electron density around the atoms and therefore differing values for the multipole parameters.

When the boundary for statistical significance is lowered to ± $2S_d$, many more multipolar parameters are statistically different from each other (Figure 5.9).

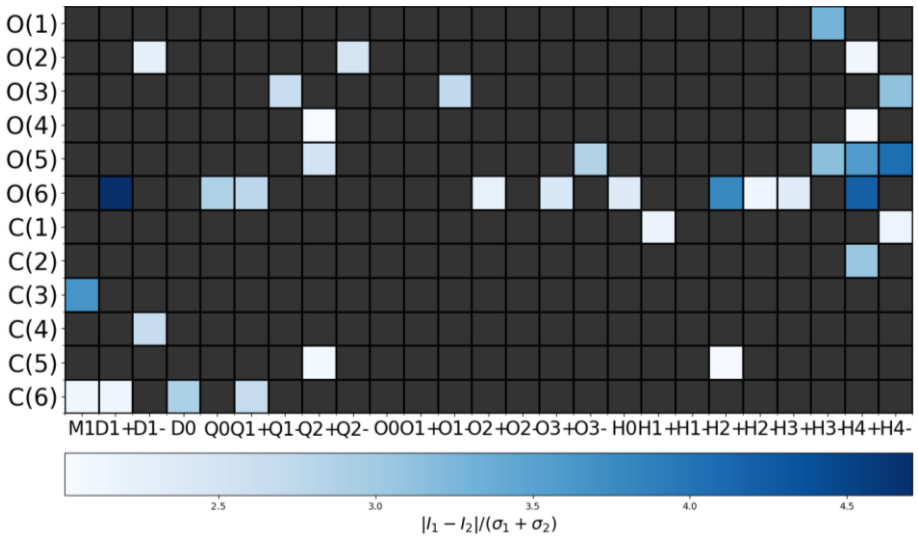


Figure 5.9 Grid indicating multipolar parameters for statistically weighted models in which ranges ± 2σ do not overlap between equivalent atoms in both molecules in L-ascorbic acid.

There are a smaller number of parameters which are statistically different when the *SHELXL* weighting scheme is applied, as shown in Figure 5.10 with a statistical significance boundary of ± $2S_d$. This is due to the increased value of σ_X for multipolar parameters where the *SHELXL* weighting has been applied (Section 3.4.4).

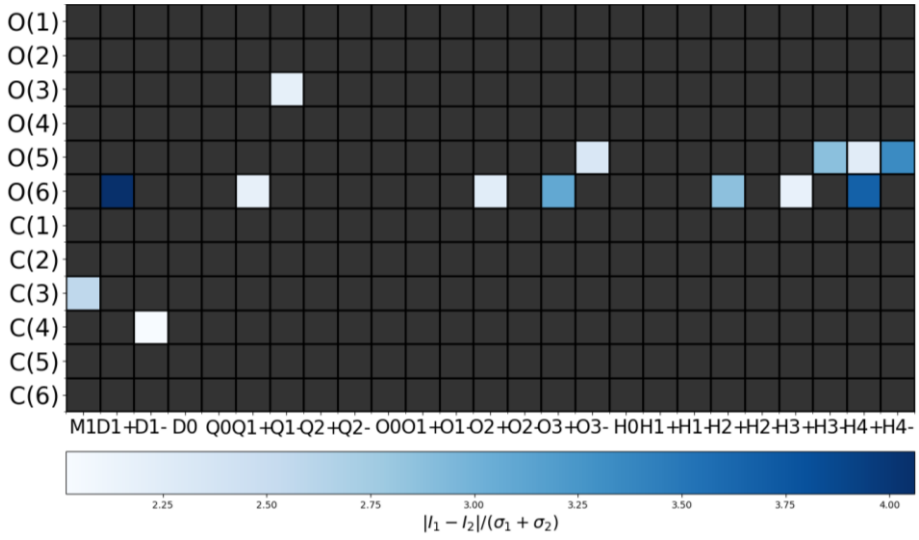


Figure 5.10 Grid indicating multipolar parameters for SHELXL weighted models in which ranges ± 2σ do not overlap between equivalent atoms in both molecules in L-ascorbic acid.

When comparing the statistical and *SHELXL* models, there was no statistical difference between the multipolar parameters of the same atoms – thereby indicating a transferability between the two models when the weighting scheme is modified. Due to this transferability, further analysis of the multipolar models and the parameters derived from them was completed using only the *SHELXL* weighted model.

5.2.3 Analysis of L-ascorbic acid multipolar model

The bond lengths reported for both molecules of L-ascorbic acid show the expected distribution, with shorter lengths calculated for those atoms which are double bonds according to Scheme 5.1 (O(1) – C(1) and C(2) – C(3)). The bonds O(5) – C(5) and O(6) – C(6), whose carbon atoms are not included in the 5 membered ring are longer than those to O(2), O(3) and O(4) which may be due to the involvement of those hydroxyl groups in the conjugation within the ring reducing the bond length. When comparing the bond lengths of the two molecules, all bond lengths except O(1) – C(1) are statistically different (Table 5.6) although there is agreement between the two molecules to ~ 2 decimal places.

Bond	Molecule 1	Molecule 2
O(1) - C(1)	1.22413(17)	1.22477(17)
O(2) - C(2)	1.35401(17)	1.35043(17)
O(3) - C(3)	1.31855(18)	1.32080(17)
O(4) - C(1)	1.35608(17)	1.35404(17)
O(4) – C(4)	1.4418(2)	1.4403(2)
C(1) - C(2)	1.45319(16)	1.45426(16)
C(2) - C(3)	1.35218(16)	1.35343(17)
C(3) - C(4)	1.49670(17)	1.49795(17)
C(4) - C(5)	1.5283(2)	1.5342(2)
C(5) - C(6)	1.52241(18)	1.52403(19)
O(5) - C(5)	1.4242(2)	1.4203(2)
O(6) – C(6)	1.4339(2)	1.4311(2)

Table 5.6 Bond lengths weighted for non-hydrogen containing bonds of multipolar model of L-ascorbic acid.

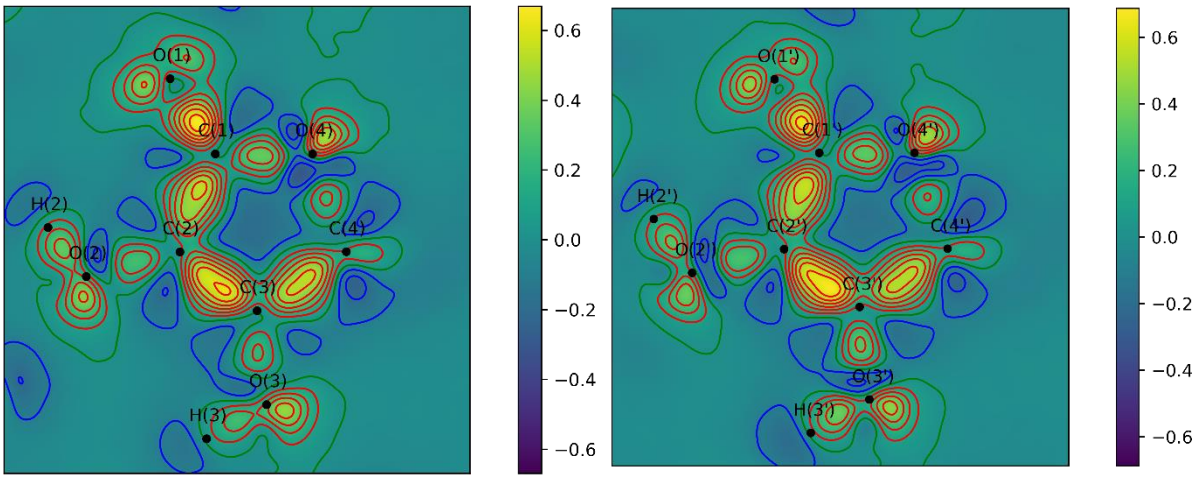


Figure 5.11 Dynamic deformation electron density of molecule 1 (left) and 2 (right) in plane of 5 membered ring. Contours are drawn at ± 0.1 eÅ^{-3}, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green.

The dynamic (Figure 5.11) and static (Figure 5.12) deformation electron densities of both molecules show similar distribution when calculated in the plane of the 5 membered ring. There appears to be little electron density within O(4) – C(4) bonds in the plane of the 5

membered rings . This is similar to that shown in the deformation densities of L-ascorbic acid in Milanesio *et al.*,[20] where the small amount of density in the C(1) – O(4) and O(4) – C(4) bonds is visible. Milanesio *et al.* did not commented on this feature. Hvolsef plotted residual electron density after the completed IAM of L-ascorbic acid in their study, showing a reduced presence of residual electron density in these bonds – when compared to the other bonds of the ring.[182,183] Therefore, it can be concluded that the multipole model recreates the same gross features as those previously published.

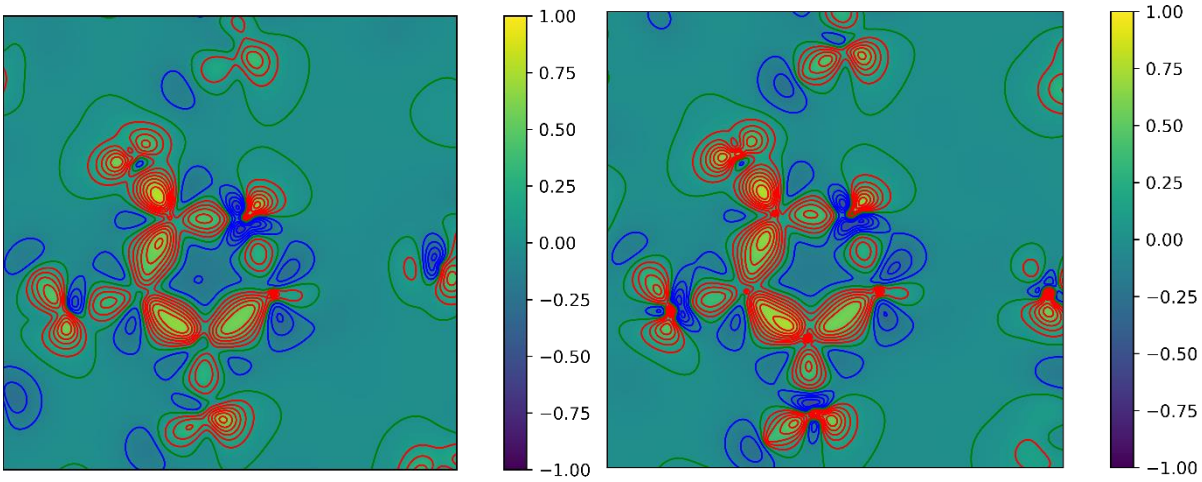
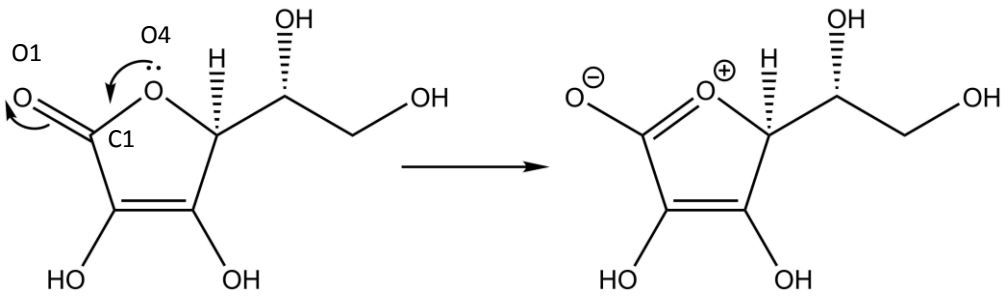


Figure 5.12 Static deformation electron density of molecule 1 (left) and 2 (right) in plane of 5 membered ring. Contours are drawn at ± 0.1 eÅ^{-3}, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

Despite the smaller amount of density within the bonds in the deformation density maps there are BCPs which occur between the pairs of atoms indicating covalent bond formation (Table 5.7) albeit with less density and charge accumulation at the O(4) – C(4) BCP than at other BCPs within the ring.



Scheme 5.2 O(4) to O(1) resonance within an L-ascorbic acid molecule.

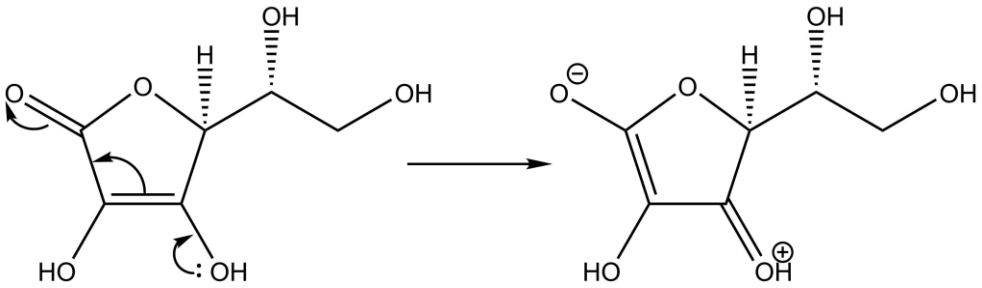
While the O(4) – C(1) bonds in both molecules also appear to have a small amount of electron density present within the static and deformation density plots, the values of ρ and $\nabla^2\rho$ at the BCP are larger than O(4) – C(4), perhaps owing to resonance between the O(4) and O(1) atoms (Scheme 5.2). The subsequent increase in double bond character, which

would occur if this was the case, is not reflected within the ellipticities of the bonds. There is little difference in the value of ellipticity for these bonds in molecule 1 while there is a larger ellipticity of O(4') – C(4') than O(4') – C(1') reported for molecule 2. However, the ellipticity at BCPs of heteropolar bonds has been suggested to be a poor measure of the π character.[22,184,185]

		Molecule 1			Molecule 2		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
O(1)	C(1)	2.948(19)	-35.911(122)	0.15	2.918(19)	-34.707(118)	0.18
O(2)	C(2)	2.124(17)	-21.925(81)	0.21	2.141(16)	-21.924(81)	0.15
O(3)	C(3)	2.241(18)	-25.374(93)	0.13	2.273(17)	-26.092(89)	0.19
O(4)	C(1)	2.184(16)	-24.583(78)	0.13	2.175(16)	-24.513(82)	0.11
O(4)	C(4)	1.732(15)	-14.891(66)	0.14	1.725(15)	-15.420(67)	0.18
C(1)	C(2)	1.917(13)	-14.966(45)	0.27	1.927(13)	-14.643(47)	0.21
C(2)	C(3)	2.266(15)	-20.301(54)	0.42	2.281(15)	-20.379(54)	0.34
C(3)	C(4)	1.863(13)	-15.815(42)	0.18	1.863(13)	-15.815(42)	0.18
O(5)	C(5)	1.886(17)	-16.566(78)	0.08	1.864(18)	-15.433(80)	0.07
O(6)	C(6)	1.781(18)	-13.233(76)	0.06	1.770(18)	-13.673(77)	0.01
C(4)	C(5)	1.769(14)	-14.807(45)	0.11	1.766(14)	-14.454(45)	0.09
C(5)	C(6)	1.803(13)	-15.365(41)	0.07	1.804(13)	-16.026(40)	0.06

Table 5.7 Properties at BCPs for non-hydrogen containing bonds for L-ascorbic acid.

The distribution of values of ρ and $\nabla^2\rho$ at BCPs are as expected for the other non-hydrogen containing bonds, with larger values for the double bonds and decreased values for O(5) – C(5), O(6) – C(6) and equivalent bonds in molecule 2. The exception to this is the C(1) – O(4) bond as described earlier. The ellipticity value at the BCP of C(2) – C(3) is more similar to that of a C – C bond in benzene (0.46)[140] than of an isolated double bond, a reduction which suggests conjugation within the ring. Conjugation is expected to occur within the ring as shown in Scheme 5.3.



Scheme 5.3 O(3) to O(1) resonance within an L-ascorbic acid molecule.

The ρ and $\nabla^2\rho$ at BCPs for intramolecular hydrogen bonds are reported in Table 5.8. All values of ρ are not statistically different between molecules as expected when the bond lengths have been constrained in the multipole model.

		Molecule 1			Molecule 2		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
O(2)	H(2)	2.221(27)	-37.661(196)	0.02	2.198(25)	-31.494(176)	0.04
O(3)	H(3)	2.151(24)	-35.846(188)	0.03	2.147(25)	-36.494(189)	0.03
O(5)	H(5)	2.278(27)	-35.117(193)	0.04	2.225(26)	-44.143(207)	0.04
O(6)	H(6)	2.264(27)	-32.314(179)	0.04	2.232(24)	-38.758(175)	0.04
C(4)	H(4)	1.845(25)	-17.836(82)	0.04	1.845(25)	-17.836(82)	0.04
C(5)	H(5A)	1.783(24)	-17.541(79)	0.03	1.783(24)	-17.541(79)	0.03
C(6)	H(6A)	1.820(27)	-18.426(87)	0.03	1.843(28)	-18.600(91)	0.07
C(6)	H(6B)	1.951(24)	-21.261(76)	0.05	1.899(24)	-20.171(72)	0.08

Table 5.8 Properties at BCPs for intramolecular hydrogen bonds in L-ascorbic acid.

BCPs were also identified for all intermolecular hydrogen bonds and detailed in Table 5.9. The length of the bond path (R_{ij}) and the length of the bond calculated in *XDGEOM* (R) are also included in the table. The R_{ij} is always longer than R, as R is calculated using as the straight-line distance and R_{ij} is calculated along the bond path which can be curved.

Atom 1	Atom 2	R	R_{ij}	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
H(2)	X2_O(5)	1.6756(2)	1.6778	0.228(16)	5.412(20)	0.06
H(3)	X1_O(1)	1.7421(1)	1.7535	0.224(10)	4.017(10)	0.04
H(5)	X2_O(6')	1.7588(2)	1.7616	0.190(12)	4.461(12)	0.05
H(6)	X2_O(5')	1.8185(2)	1.8232	0.165(12)	3.937(10)	0.04
H(2')	O(6)	1.6471(1)	1.6500	0.295(14)	5.289(18)	0.07
H(3')	X1_O(1')	1.7316(1)	1.741	0.207(11)	4.161(11)	0.05
H(5')	X2_O(2')	1.8424(2)	1.8717	0.127(10)	3.365(8)	0.09
H(6')	X1_O(2)	2.0418(1)	2.0895	0.094(5)	1.979(1)	0.12

Table 5.9 Properties at BCPs for intermolecular hydrogen bonds for L-ascorbic acid.

In general, the bonds with shorter lengths had a larger value for both ρ and $\nabla^2\rho$. The longest hydrogen bond H(6') to O(2) has the smallest ρ at 0.099, but this value is still very statistically significant, owing to a low value of standard uncertainty associated with the value.

Atom 1	**Atom 2**	**Milanesio *et al.*[20]**	**Thesis**
H(3)	X1_O(1)	0.240(24)	0.224(10)
H(2')	O(6)	0.410(5)	0.295(14)
H(3')	X1_O(1')	0.2000(24)	0.295(14)
H(6')	X1_O(2)	0.0900(12)	0.094(5)

Table 5.10 Reported ρ at BCP for selected intermolecular hydrogen bonds within L-ascorbic acid structure by Milanesio et al. X indicates which symmetry equivalent molecule the hydrogen interacts with.

Milanesio *et al.* report ρ at the BCPs for four of the eight hydrogen bonds identified. These are listed in Table 5.10, along with the corresponding values calculated in this thesis. Three of the values are not statistically different from each other, suggesting the reproducibility of the study against that of previous results. However, the value of the density for H(2') – O(6) is much larger for the Milanesio study. This may be because the Milanesio study took place

at a higher temperature which may have altered the distance and angle between the two bonds affecting the calculated value or due to differences in the defined axes for the multipolar parameters.

5.3 Conclusions

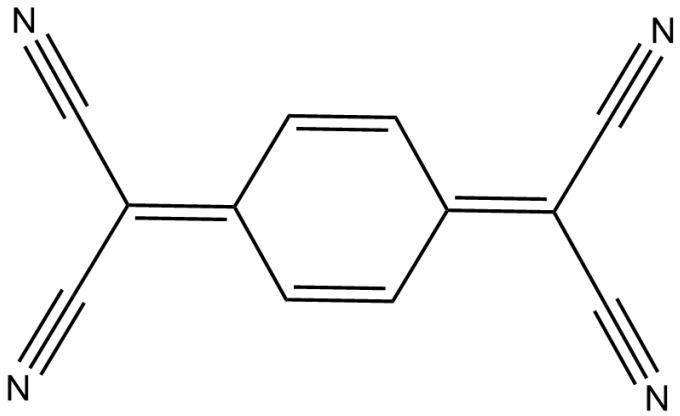
The multipolar model in this thesis displays a good agreement with the previous study, reported by Milanesio *et al*. Additionally, the comparison of multipole parameters for the two molecules in the unit cell showed a good degree of transferability. It would be expected that differences in multipolar parameters between the same atom in the different molecules will not be statistically significant, as they are in similar chemical environments and the conformational difference is very small. This was found to be the case for most multipolar parameters. As the local coordinate system of the atoms within the refinement have been defined by vectors between atoms, the differences in atomic positions relative to each other within the two molecules may be the cause of the different values for multipoles of O(5) and O(6). The differences shown may be due to this particular data processing strategy and are seen for both the statistical and *SHELXL* weighted structures.

It is evident from this chapter that white radiation contamination is a more prevalent problem than previously expected when collecting crystallographic data from small molecule crystals with this particular experimental setup. Contamination may be hidden by including fast scans in the calculation of the precession images used to check for the presence of this effect. The filter of 100 µm used for this data collection was not thick enough to filter all the white radiation and therefore a filter of 150 µm, as was used in other collections in this thesis, is preferable.

CAPOW has proven an extremely useful tool for the identification of systematic errors within the multipolar refinement. If the optimised *SHELXL* weighting scheme does not return a distribution of residuals between ~ ± 4 then additional systematic errors may still be present in the data. These reflections must be removed from the data, or the data recollected to remove the effect of any systematic errors, such white radiation contamination, before multipolar analysis can be relied upon. This also highlights the importance of assessing the whole range of residuals within a normal probability plot, as opposed to just those between ± 4, to enable the identification of such outliers.

Chapter 6. Investigation of conjugation using charge density

Tetracyanoquinodimethanes are small molecules with a high degree of conjugation. They are therefore ideal candidates to investigate the effect of changes in the conjugation on the distribution of electron density within a molecule. Charge density studies of two similar molecules, 7,7,8,8-tetracyanoquinodimethane (TCNQ) and 1,4-bis(dicyanomethylene)cyclohexane (DCMC), are presented and compared. Both molecules contain two malononitrile substituents that are oriented in a para position to each other on a 6 membered ring. The bond order, i.e. the conjugation present, within the ring is different between the two molecules.

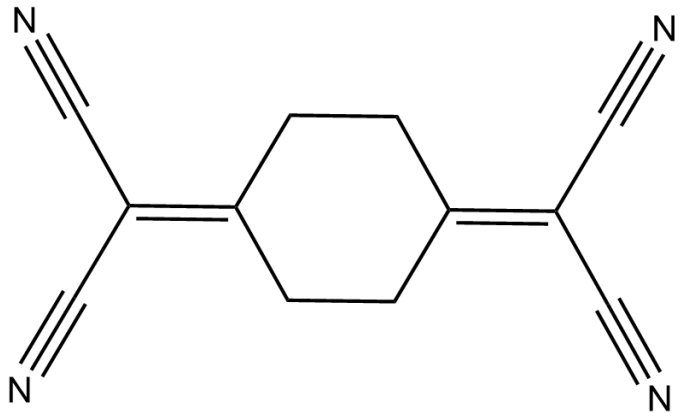


Scheme 6.1 7,7,8,8-tetracyanoquinodimethane (TCNQ).

TCNQ (Scheme 6.1), as previously described in Chapters 2 and 3, is a planar molecule with alternating double and single bonds allowing delocalisation of electron density throughout the molecule. It was first characterised by single crystal X-ray diffraction in 1965[186] and is well-known for co-crystallising with tetrathiafulvalene (TTF) to act as the electron acceptor in a charge transfer complex (TTF-TCNQ).[187,188] The ability to delocalise electron density throughout the molecule also allows the formation of a stable negative TCNQ ion. To date an experimentally derived multipole model has not been published for TCNQ alone, although it is known to have been studied.[137]

DCMC is a partially saturated analogue of TCNQ (Scheme 6.2), where the double bonds within the ring have all been hydrogenated. It can be obtained as an intermediate in the

process of synthesising TCNQ.[189] The crystal structure of which was first reported by Iwasaki and Aihara.[190] Due to its saturated cyclohexane ring, DCMC is not a planar molecule and instead typically exhibits the chair conformation.



Scheme 6.2 1,4-bis(dicyanomethylene)cyclohexane (DCMC).

6.1 Methods

TCNQ (98%) was used as purchased from Alfa Aesar without further purification. It was recrystallised by slow cooling of a saturated solution of TCNQ and acetonitrile (purchased from Fischer Scientific, HPLC Gradient grade) to produce yellow rectangular crystals suitable for single crystal X-ray diffraction studies. DCMC was provided by Dr J. Sellars (Newcastle University) and was synthesised using the method described in Section 10.4.1. Crystals of DCMC were obtained by recrystallisation via slow cooling of a saturated solution of acetonitrile, producing colourless prisms.

In the data collection for both crystals, a single crystal was mounted on a *MiTeGen* loop and slow cooled to 100K at a rate of 1 K per minute using an Oxford Cryosystems N2 cryostream. X-rays were generated using an Incoatec IμS 3.0 Ag source (λ = 0.56086 Å) with a 150 μm aluminium filter included. Data collection strategies were calculated in the *APEX3* GUI to achieve 99% completeness to a resolution of 0.45 Å and a multiplicity of at least 12 for 90% of reflections within that range. The full experimental run list of the data collection for each study is listed in Section 10.4.2. The diffraction patterns were measured on a Photon II CPAD detector using shutterless operation at a distance of 65 mm.

The collected data were integrated in *SAINT* using default integration algorithm, with the best-plane background to a resolution of 0.48 Å. The data were scaled and merged in *SADABS* using the default error model. A numerical absorption correction was not applied in TCNQ as the crystal was not excessively anisotropic (0.152 x 0.224 x 0.306 mm^3) and it was shown in Chapter 3 that the correction made little difference to the final multipole model.

The DCMC multipolar model, however, required a numerical absorption correction to be applied to allow convergence of the multipolar refinement (as all multipolar parameter lists trialled did not converge when the correction was not applied). The crystal selected was a rectangular prism (0.190 x 0.246 x 0.672 mm^3) which would result in a larger path length difference for different reflections necessitating the correction. Due to the larger size of the crystal, a correction for the overloaded reflections using a fast scan was also applied to DCMC. 17 overloaded reflections were corrected for DCMC.

Plots of the binned R_{int} with respect to resolution for TCNQ and DCMC for the unmerged version of the *xd.hkl* file show low values for R_{int} at low angle < 0.03 %, rising to a maximum of ~ 0.18 % at 0.48 Å (Figure 6.1). There is a pronounced bump in the R_{int} of DCMC at ~ 0.6 Å resolution. This is an artefact of merging data measured with different exposure times.

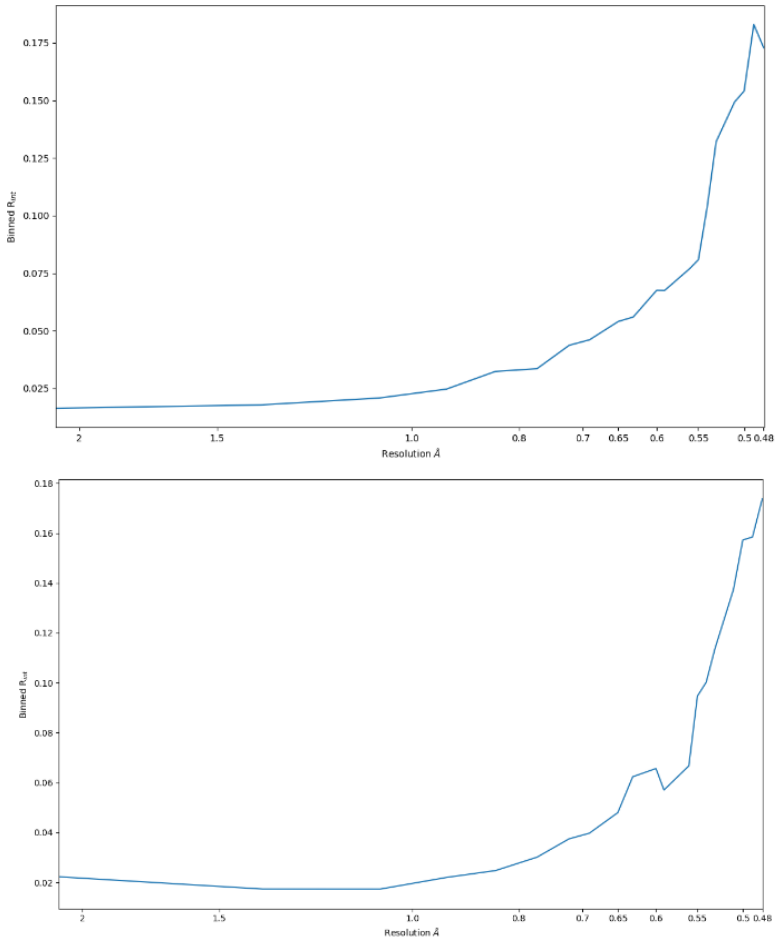


Figure 6.1 Binned R_{int} plot for reflections unmerged version of xd.hkl file for TCNQ and DCMC.

Space groups were determined in *XPREP* and the spherical atom model solution and the refinements were performed in the *OLEX2* GUI using *XT* and *XL* respectively. A summary of the spherical atom refinements is given in Table 6.1. 4

	TCNQ	DCMC
Formula	$C_{12}H_4N_4$	$C_{12}H_8N_4$
Space group	*C2/c*	*P2₁/n*
Crystal system	Monoclinic	Monoclinic
a (Å)	8.8626(8)	7.0798(7)
b (Å)	6.8898(7)	7.0375(6)
c (Å)	16.3890(16)	10.5042(10)
β (°)	98.220(3)	91.531(2)
V (Å^3)	990.46(17)	523.18(8)
ρ (g/cm^3)	1.369	1.322
Z	4	2
Z'	0.5	0.5

Table 6.1 Unit cell parameters of TCNQ and DCMC calculated with spherical atom model.

Global refinement statistics calculated after the spherical atom treatment of both molecules are reported in Table 6.2. Both datasets show the presence of peaks of residual electron density, that are not accounted for by the spherical atom model, with a maximum of ~ 0.8 eÅ^{-3} for each structure – indicating a large range of residual electron density which resides principally between atoms where covalent bonding density would be expected. This could therefore be modelled using multipolar parameters.

	TCNQ	DCMC
R_1 (%)	3.45	3.79
wR_2 (%)	11.12	12.26
R_{int} (%)	4.13	3.55
Average multiplicity	20.00	12.16
GooF	1.134	1.093
Total reflections	94763	60655
Independent reflections	4738	4990
Mean I/σ	75.8321	55.2493
e hole (eÅ^{-3})	-0.225	-0.192
e peak (eÅ^{-3})	0.868	0.882
Weighting scheme	*SHELXL* a = 0.060 b = 0.150	*SHELXL* a = 0.071 b = 0.034

Table 6.2 Refinement statistics using all collected data from spherical atom model.

Multipolar refinement was performed using *XD* with the least-squares minimisation calculated against F^2. As far as was possible the local coordinate system of each atom was kept consistent between the two models in order to facilitate the comparison of multipolar parameters, with the atoms labelled as in Figure 6.2. The values for anisotropic ADPs for hydrogen atoms were estimated using the *SHADE* server[130] and the bond lengths fixed to those of expected lengths from neutron diffraction studies. Hydrogen positions were then

allowed to refine using only low angle reflections (0 – 0.5 $Å^{-1}$) before the bond lengths were fixed at expected hydrogen bond lengths derived from neutron refinements.

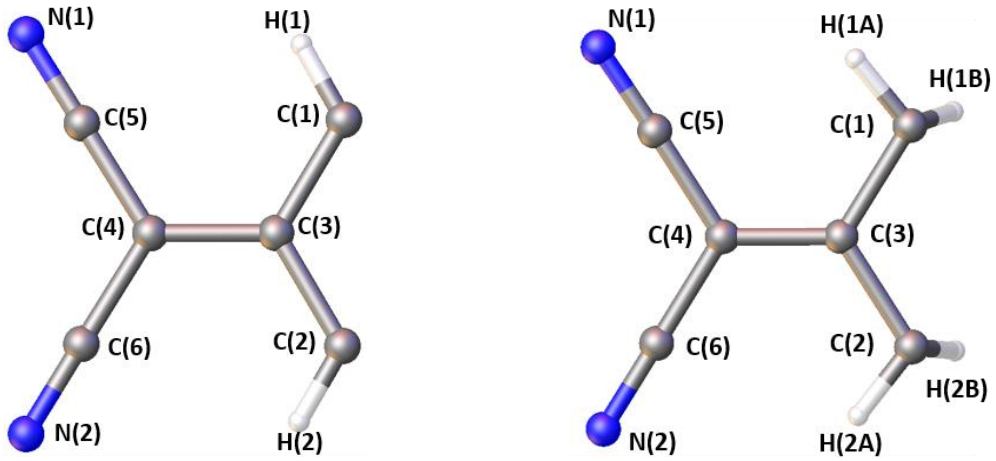


Figure 6.2 Atom labels for TCNQ and DCMC, as used in multipolar refinements.

Multipoles up to the hexadecapole level were included for non-hydrogen atoms, while the monopole and one bond directed dipole were refined for hydrogen atoms. κ and κ' parameters were then included in the refinement. The full refinement strategy is included in Section 10.4.3. Once the multipole parameters had reached convergence at the end of the refinement strategy, the hydrogen U_{ij} parameters were recalculated using the *SHADE* server and further iterations of refinement were performed after inputting the new values. This move was found to be an unimportant step as U_{ij} values calculated for hydrogen atoms using the converged multipolar refinement did not change appreciably.

After convergence of the refinement optimal values for the *a* and *b* parameters of a *SHELXL* weighting scheme were calculated using *CAPOW* for both molecules. The weighting parameter values were optimised as described in Chapter 3 until convergence of the values occurred to 4 significant figures. The optimised value for the *a* and *b* parameters for TCNQ were *a* = 0.0058 and *b* = 0.0219, whereas for DCMC they were *a* = 0.0121 and *b* = 0.0035.

6.2 Verification of multipole models

The refinement statistics from the completed multipole refinement are detailed in Table 6.3. Both molecules see a reduction in the R_1 value and range of residual density when compared to that of the IAM (Table 6.2) signifying that the addition of multipolar parameters results in a better agreement between calculated and observed structure factors. The application of a weighting scheme resulted in a reduction in the range of residual electron density, e_{gross} and wGooF along with an increase in $d^f(0)$, which implies that the addition of the *SHELXL*

weighting scheme produced a more flat and featureless distribution of residual electron density (as seen in previous chapters).

	TCNQ		DCMC	
Weighting	Statistical	*SHELXL*	Statistical	*SHELXL*
Total reflections	4469	4469	4771	4771
Reflections used	3944	3944	4138	4138
R_2 (%)	1.49	1.49	2.13	2.10
R_1 (%)	1.29	1.28	1.41	1.40
wR_2 (%)	2.0	2.32	2.31	2.88
wGooF	1.4483	1.0968	1.4972	1.0844
$d^f(0)$	2.4992	2.4995	2.4331	2.4345
e_{gross}	1.2725	1.2582	1.6042	1.5798
e_{net}	0.0501	0.0544	-0.0601	-0.0408
e hole ($eÅ^{-3}$)	0.110	0.108	0.131	0.128
e peak ($eÅ^{-3}$)	-0.095	-0.090	-0.097	-0.096
N_{ref}/N_{rev}	14.0857	14.0857	15.5704	14.5704

Table 6.3 Multipolar refinement statistics for statistically and SHELXL weighted TCNQ and DCMC.

Multipolar refinements of both TCNQ and DCMC were further verified using normal probability plots, residual density analysis and residual electron density plots. Both refinements did not return a normal distribution of residuals when the statistical weighting scheme was used according to the normal probability plots (Figure 6.3), with the shape of the plot suggesting a slight underestimation of standard uncertainties. The application of optimised parameters for the *SHELXL* weighting scheme resulted in a more normal distribution of residual values with a range of ~ ± 4 (Figure 6.4). These plots do not suggest the presence of any white radiation contamination (nor do precession images which are calculated from the data).

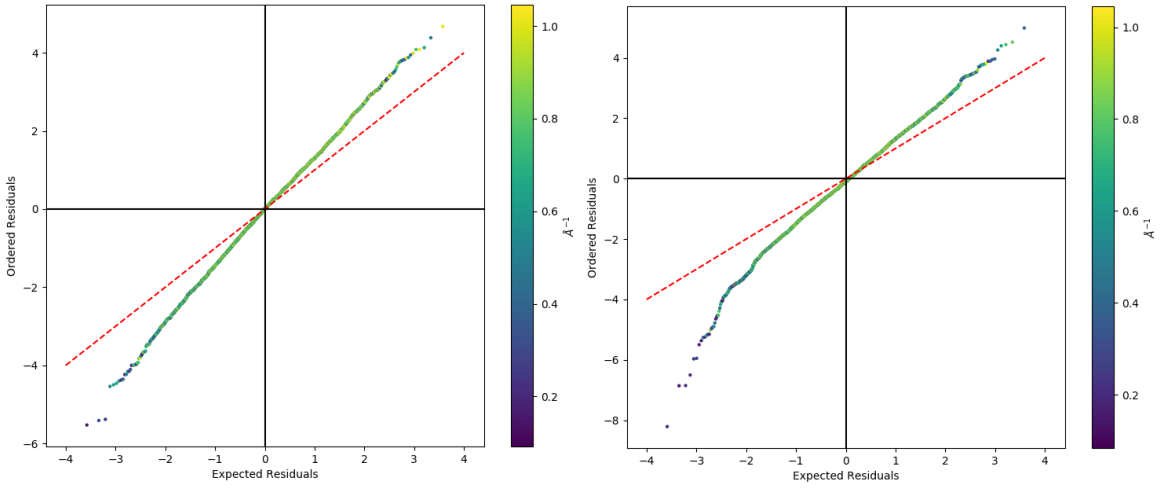


Figure 6.3 Normal probability plot for statistically weighted model of TCNQ (left) and DCMC (right).

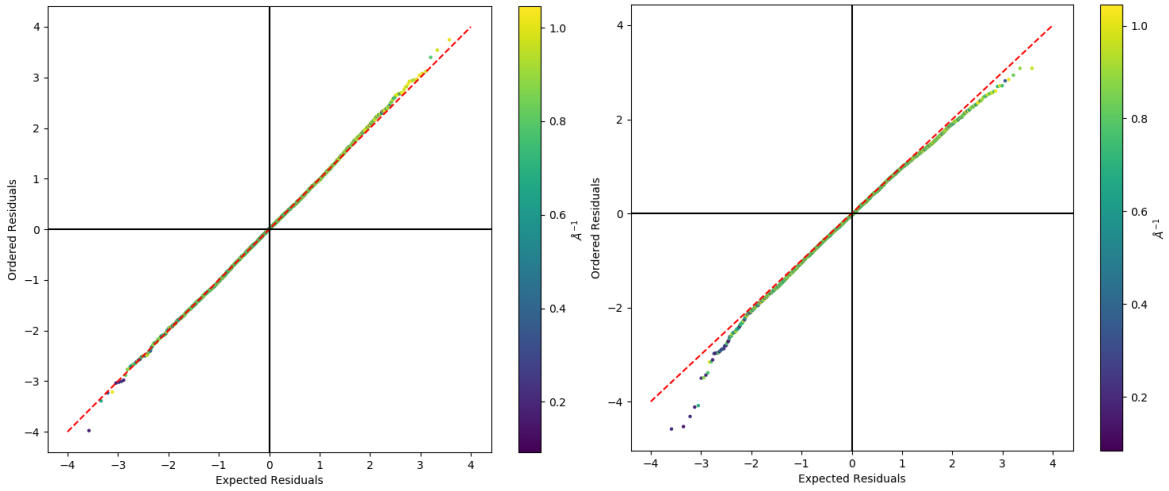


Figure 6.4 Normal probability plots for SHELXL weighted multipole model of TCNQ (left) and DCMC (right).

Residual electron density plots within the plane of the asymmetric unit show approximately the same distribution of residuals for statistical (Figure 10.40) and *SHELXL* (Figure 6.5) weighted models for both molecules. The residual density within the plane of the molecule is relatively flat with few peaks or troughs indicating a good description of observed electron density within these planes.

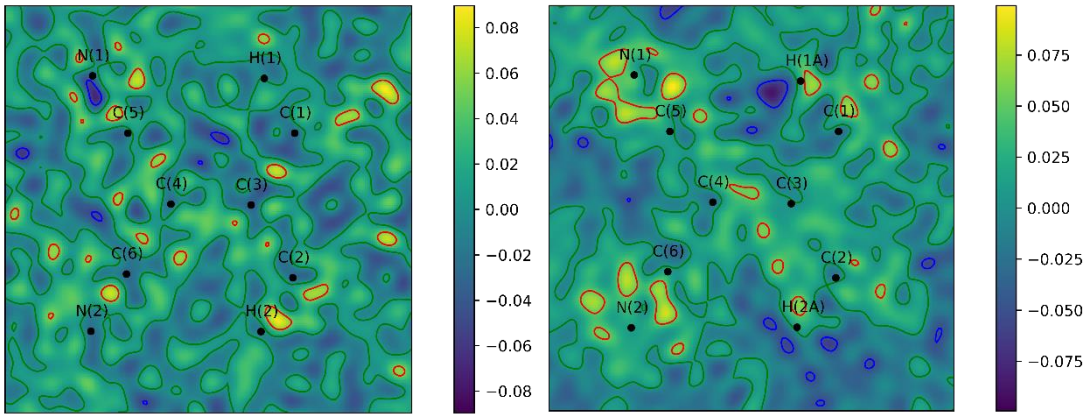


Figure 6.5 Residual electron density plot of SHELXL weighted TCNQ and DCMC within the plane of the asymmetric unit. Contours are drawn at ± 0.05 eÅ^{-3} levels, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green.

Fractal dimensionality plots also showed the same relatively featureless distribution of residuals for a 3D grid of residual electron density calculated to enclose the asymmetric unit. Both plots for the *SHELXL* weighted model (Figure 6.6) show an uneven distribution for positive and negative ρ; however, there are no large features in either plot which would indicate the presence of systematic errors. The fractal dimensionality plots for the statistically weighted models (Figure 10.41) also show the same distribution of values. The difference in mean square displacement amplitude along the bond paths of all non-hydrogen atoms were all < 1×10^{-3}, indicating that all non-hydrogen bonds pass the Hirshfeld rigid-bond test.

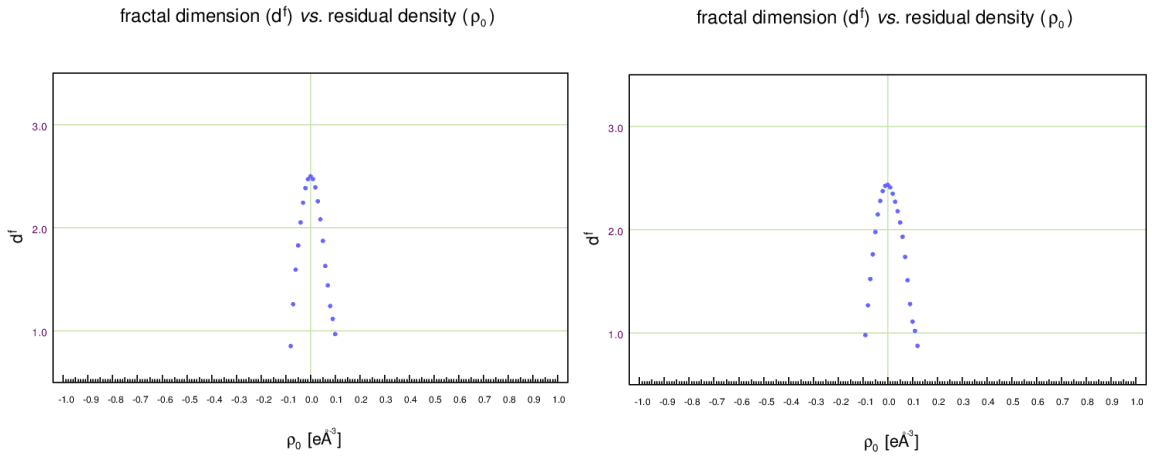


Figure 6.6 Fractal dimensionality plots of SHELXL weighted TCNQ (left) and DCMC (right).

These statistics suggest that the multipole models for both statistically and *SHELXL* weighted TCNQ and DCMC describe the data well and do not suggest the presence of any strong systematic errors.

6.3 Comparison of TCNQ and DCMC

There are no statistical differences between the statistical and *SHELXL* weighted multipolar parameters within each molecule, therefore, comparison of the multipolar parameters will be performed using *SHELXL* weighted refinements only (due to the improved refinement statistics). A number of properties derived from the multipolar analysis of TCNQ and DCMC are compared to investigate the differences between the electron distribution in the molecules.

Atom 1	Atom 2	TCNQ	DCMC
N(1)	C(5)	1.1570(2)	1.1539(2)
N(2)	C(6)	1.1570(2)	1.1562(2)
C(1)	C(3)	1.44375(18)	1.4928(2)
C(2)	C(3)	1.44376(18)	1.49423(19)
C(1)	C(2*)	1.35521(19)	1.5404(2)
C(3)	C(4)	1.38037(19)	1.35445(19)
C(4)	C(5)	1.4250(2)	1.4295(2)
C(4)	C(6)	1.4253(2)	1.4302(2)

Table 6.4 Bond lengths of non-hydrogen bonds within TCNQ and DCMC (Å).

Within the bond lengths derived from the model (Table 6.4) it is possible to see the alternating double and single bond arrangement pattern of C – C bond lengths indicating a quinoidal molecule in the structure of TCNQ. This is not present in DCMC, which is not a quinoidal molecule. The bond length for C(1) – C(2*) in both molecules reflects the difference in the bond order within the ring, with DCMC having a significantly longer bond.

The bond lengths in DCMC are not statistically different to those of the reported structure by Iwasaki and Aihra,[190] apart from C(5) – N(1) (1.134(4) Å). The differing bond lengths between C(1) – C(3) and C(2) – C(3) vs C(1) – C(2*) also mirror those of the published results. The length of C(1) – C(2*) is much closer to that of the C – C bond lengths in structures of cyclohexane, as would be expected due to the saturation of the ring in DCMC.[191] The variation in bond lengths within the ring can be explained by assuming that C(1) – C(3) and C(2) – C(3) are sp^3-sp^2 hybridised (resulting in shorter bond lengths), whereas C(1) – C(2*) is sp^3 – sp^3 hybridised.[190,192]

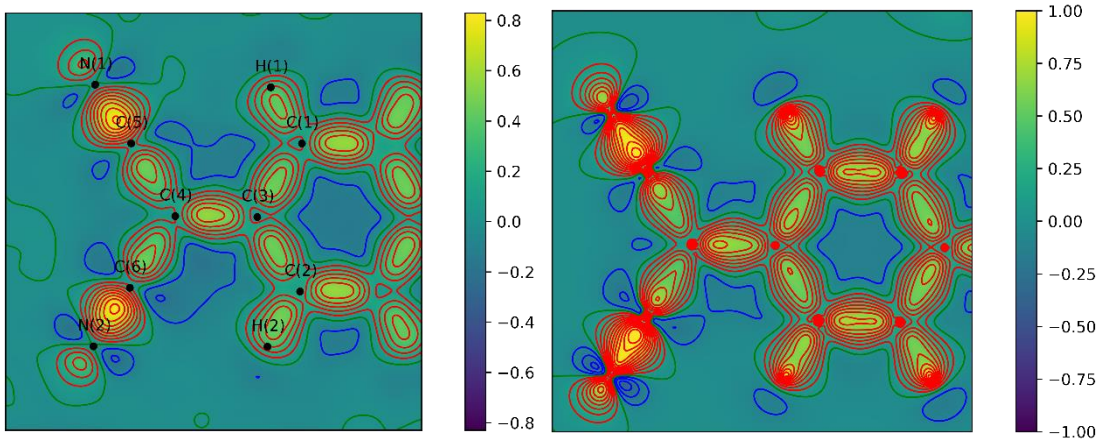


Figure 6.7 Dynamic and static deformation electron density plots for TCNQ. Contours are drawn at ± 0.1 eÅ^{-3} levels, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green. The colourbar for the static deformation density plot is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

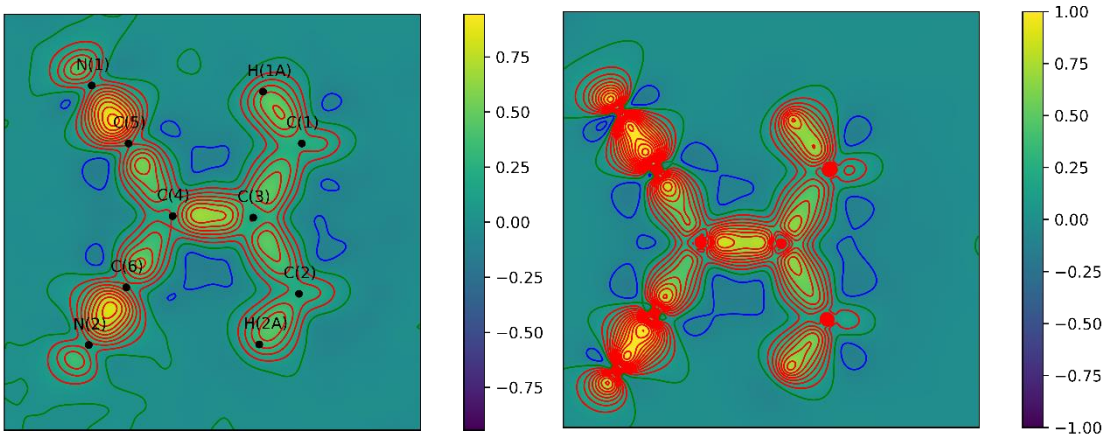


Figure 6.8 Dynamic and static deformation density plots for DCMC. Contours are drawn at ± 0.1 eÅ^{-3} levels, where red indicates positive contour and blue indicates negative contours. The zero contour is present in green. The colourbar for the static deformation density plot cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

The dynamic and static deformation electron density plots show similar features in both TCNQ (Figure 6.7) and DCMC (Figure 6.8). The C(1) – C(2*) bond is not visible for DCMC in this orientation due to the molecule not being planar.

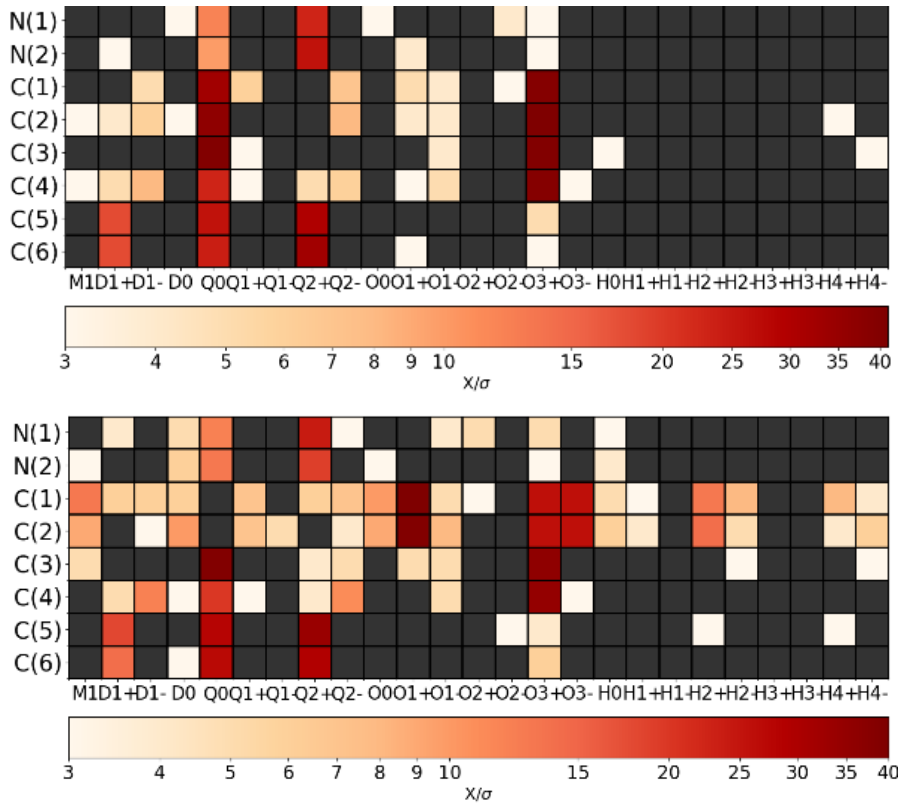


Figure 6.9 Statistical significance of multipolar parameters (with respect to 0) for TCNQ (upper) and DCMC (lower).

Both TCNQ and DCMC have statistically significant values for multipolar parameters up to hexadecapolar level, though to less of an extent for TCNQ (Figure 6.9). The main difference in significance of parameters is for the C(1) and C(2) atoms as many multipolar parameters (including 6 hexadecapolar parameters) have a statistically significant populations in DCMC. This could be due to the change in geometry of the two atoms. In DCMC there are 4 different bonds to C(1) and C(2) atoms in a tetrahedral geometry, while in TCNQ bonds are only within the plane of the molecule (Figure 6.10). This means that as the direction of the bonds is different for C(1) and C(2) in the two molecueles, the electron distribution surrounding the atoms would also be different to allow this bonding to occur. This could result in more multipolar parameters having signficiant populations in order to model this density which now involves atoms in a 3D distribution (tetrahedral) instead of 2D (planar).

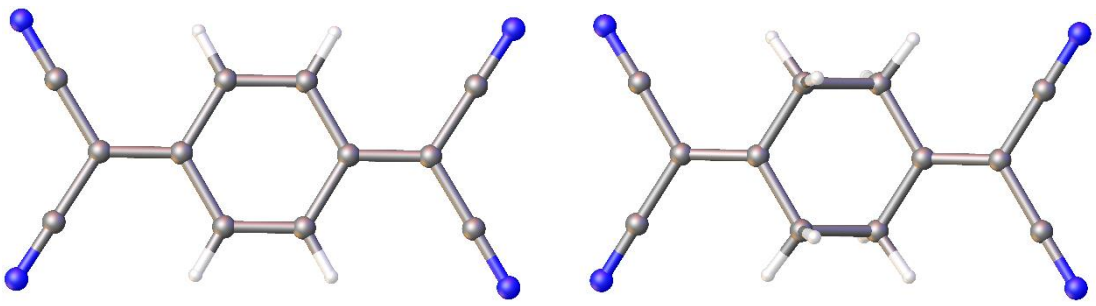


Figure 6.10 Representation of TCNQ and DCMC molecules created in Olex2.

There are a number of differences in the multipole parameters between the two molecules, as illustrated in Figure 6.11, where statistically different multipole populations between TCNQ and DCMC are coloured with respect to the S_d. Parameters in C(1) and C(2) are shown to have large statsitical differences in the values of population of multipolar parameters, up to a maximum of ~ 35 S_d. This may be expected when considering the significance plots of the two molecules (Figure 6.9) and the difference in the chemical environments of the two atoms between TCNQ and DCMC. Most of the populations of multipolar parameters for atoms in more equivalent environments are not statistically different from each other except for a few of the multipolar parameters of N(1) and N(2), where the difference between the values is only slightly greater than 3 S_d. This suggests that overall there is transferability in the multipolar parameters of atoms in chemically similar environments between the two molecules.

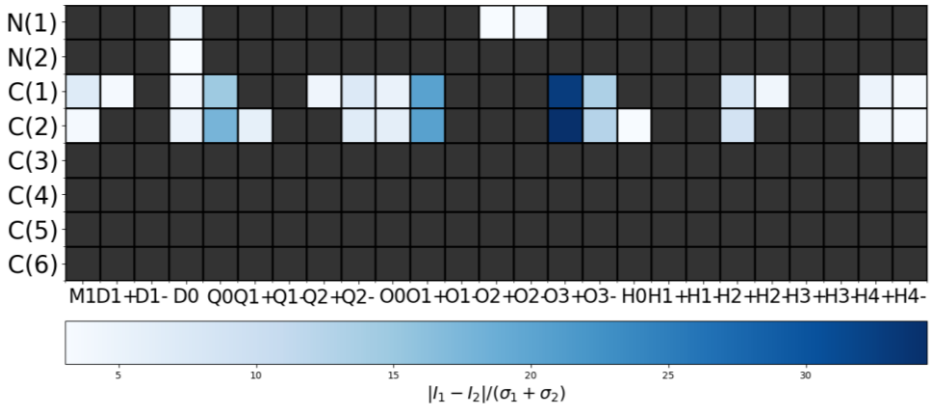


Figure 6.11 Plot showing the statistically different multipole parameters of TCNQ and DCMC illustrated in blue. The colour scheme relates to the number of total standard deviations of the parameters the values are apart. Squares in grey correspond to no statistical difference between the multipolar value in TCNQ and DCMC.

BCPs were calculated for intramolecular distances (0.8 – 1.6 Å) using *XDPROP* and values of ρ, $\nabla^2\rho$ and ellipticity at BCPs are reported for all non-hydrogen bonds in Table 6.5. For TCNQ and DCMC, the values of ρ and $\nabla^2\rho$ at the BCPs decrease based on the expected bond order.

When comparing BCPs in similar environments between TCNQ and DCMC, the values for N(2) – C(6), C(4) – C(5) and C(4) – C(6) show no statistical difference in ρ between the two molecules, which is expected due to the bonds being approximately similar. However, N(1) – C(5) has a greater value of ρ and $\nabla^2\rho$ in DCMC, perhaps due to the shorter bond length. The C(3) – C(4) BCP has a larger density and charge accumulation in DCMC, suggesting less conjugation involving this bond, also represented by the higher ellipticity at the DCMC BCP. However, this value is still lower than an isolated double bond, and closer to that of benzene,[140] implying a degree of conjugation may still exist involving this bond (perhaps involving the malononitrile groups).

		TCNQ			DCMC		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
N(1)	C(5)	3.448(18)	-36.997(135)	0.05	3.612(16)	-41.401(117)	0.02
N(2)	C(6)	3.498(18)	-38.748(135)	0.03	3.482(17)	-34.793(130)	0.06
C(1)	C(3)	1.970(9)	-16.212(32)	0.15	1.792(8)	-12.491(28)	0.04
C(2)	C(3)	1.946(9)	-15.992(31)	0.12	1.828(8)	-14.266(27)	0.04
C(1)	C(2*)	2.292(19)	-20.229(53)	0.24	1.612(12)	-10.766(30)	0.05
C(3)	C(4)	2.191(10)	-18.902(37)	0.24	2.344(10)	-19.536(39)	0.27
C(4)	C(5)	1.915(9)	-13.948(32)	0.11	1.931(9)	-12.936(34)	0.09
C(4)	C(6)	1.938(9)	-14.378(33)	0.15	1.889(10)	-12.540(35)	0.09

Table 6.5 Selected intramolecular BCPs for TCNQ and DCMC

The overall values of ellipticity at the BCPs differs between the two molecules. The bonds that are expected to be single in TCNQ all have an elevated ellipticity, which would suggest a degree of conjugation between all the atoms. In DCMC, the cyclohexane ring bonds have a lower ellipticities suggesting negligible conjugation, which would be expected as lack of a double bond within the ring would prevent conjugation within the ring. In the malononitrile group, the ellipticity of the C – C bonds is slightly raised suggesting some conjugation remaining in the bonds after reduction of TCNQ to DCMC.

There is no statistical difference in the value of net atomic charges between chemically equivalent atoms in TCNQ and DCMC (Table 6.6 and Table 6.7). All hydrogen atoms within both molecules are positively charged. This could be expected due to the electronegativity difference between C and H and the lack of core electrons in H resulting in the single H electron being drawn closer to carbon, away from the hydrogen nucleus, leaving the atom partially positively charged. The C(1) and C(2) atoms show a large increase in net atomic charge in DCMC, which may be expected as the number of hydrogens bonded to the carbon has increased, therefore a greater amount of electron density can be drawn towards those atoms. The extra charge on C(1) and C(2) could in turn have resulted in the statistical difference of the net atomic charge of C(3) as the charge can increase to balance this electrostatically.

Atom	P	κ	κ'	Net charge
N(1)	5.053(20)	1.000(1)	0.992	-0.053(20)
N(2)	5.044(20)	1.000(1)	0.992	-0.043(20)
C(1)	4.011(17)	1.015(2)	0.982	-0.011(17)
C(2)	4.066(19)	1.015(2)	0.982	-0.065(19)
C(3)	3.963(20)	1.013(2)	0.978	+0.037(20)
C(4)	4.064(20)	1.013(2)	0.978	-0.064(20)
C(5)	4.054(25)	1.030(2)	0.985	-0.054(25)
C(6)	4.042(24)	1.030(2)	0.985	-0.041(24)
H(1)	0.864(8)	1.2	1.2	+0.135(8)
H(2)	0.838(8)	1.2	1.2	+0.162(8)

Table 6.6 Net atomic charges for TCNQ.

Atom	P	κ	κ'	Net Charge
N(1)	5.042(23)	1.013(2)	1.014	-0.042(23)
N(2)	5.095(24)	1.013(2)	1.014	-0.094(24)
C(1)	4.271(21)	1.016(2)	0.933	-0.270(21)
C(2)	4.204(22)	1.016(2)	0.933	-0.204(22)
C(3)	3.884(21)	1.037(2)	0.97	+0.116(21)
C(4)	4.036(22)	1.037(2)	0.97	-0.036(22)
C(5)	4.067(25)	1.051(2)	0.975	-0.067(25)
C(6)	4.021(27)	1.051(2)	0.975	-0.020(27)
H(1A)	0.872(10)	1.2	1.2	+0.127(10)
H(1B)	0.823(10)	1.2	1.2	+0.177(10)
H(2A)	0.875(10)	1.2	1.2	+0.124(10)
H(2B)	0.809(10)	1.2	1.2	+0.191(10)

Table 6.7 Net atomic charges for DCMC.

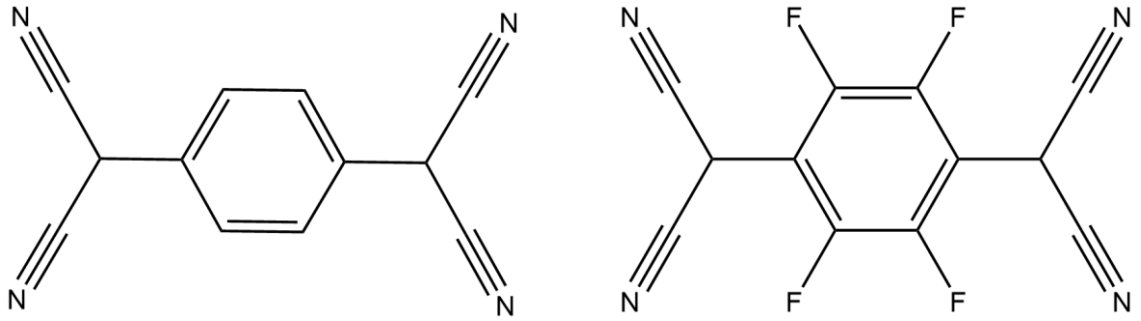
6.4 Conclusions

This study has shown that there is general transferability of multipolar parameters for TCNQ and DCMC for atoms in chemically equivalent environments. However, when the geometry of the bonding changes (e.g. C(1) and C(2)) the resulting multipolar parameters are broadly statistically different. The transferability of the multipolar parameters for crystallographically equivalent atoms within molecules that do contain changes in bonding is very encouraging, especially in relation to the applicability of using multipolar values from TAAM databanks.

When comparing the multipole models of the two structures, the ellipticities at bond critical points for DCMC suggest minimal conjugation within the molecule – apart from within the malononitrile group. The ellipticities for all bonds in TCNQ are higher, indicating a greater degree of conjugation – which extends across the whole molecule. The difference in values at BCPs and net atomic charges can all be explained by the change in bonding and hybridisation of atoms in the atoms of the two molecules. These results clearly demonstrate

the validity of the use of charge density studies to investigate fundamental chemical features.

Due to time limitations, further experiments using other similar molecules were not undertaken. The compounds TCNX and the equivalent F_4TCNX (Scheme 6.3), reductions of TCNQ and F_4TCNQ respectively, would be of interest to enable the study of a wider range of comparable compounds with differences in conjugation. These compounds are intermediates in the formation of TCNQ[189] and F_4TCNQ[193] and have been identified within scientific literature but have yet to be structurally characterised by crystallography. The change in geometry at the C(4) atom due to the addition of hydrogen atoms in these compounds would likely result in differences in the multipolar parameters. Also, the bond length and ellipticities of the C(1) – C(3) and C(2) – C(3) bonds would be expected to increase due to the formation of an aromatic ring. This would likely result in statistical differences within a number of multipolar parameters when compared to TCNQ and F_4TCNQ.



Scheme 6.3 TCNX and F_4TCNX.

Chapter 7. Diffraction frame format conversion

Diffraction frames are the files created by X-ray detectors during an experiment and contain the intensities of the measured diffraction pattern along with experimental metadata. These files are output from detectors in a variety of different formats that usually depend on the instrument, with many manufacturers using an arbitrary 'standard' format across their product range. The way in which the information in the frame is structured depends upon the file format specification – a document which describes the layout of the file; the fields that are expected, the units and number of decimal places that variables are rounded to, as well as what encoding data are required to be stored in (e.g. ASCII, binary, hexadecimal, etc.). As well as there being a range of diffraction frame formats, there are also a number of different diffraction integration software packages: both commercial and non-profit. The specific frame formats supported by a particular integration software often vary, which can lead to data collected using certain instruments not being compatible with a particular program.

In 2016, beamline I19 at Diamond Light Source (a third generation synchrotron located in Harwell, UK) underwent an upgrade to one its of diffractometers (Instrument EH1), which included exchanging its Rigaku Saturn 724+ CCD detector for a Dectris Pilatus 2M HPAD.[194] The new detector allow 'standard' data collections to be performed on a much shorter timescale owing to the shutterless readout capability of the 2M (reducing the collection time from 40-60 mins to 15 mins).[195] Diffraction frames from the 2M detector are output in *.cbf* format,[196] a format not compatible with some industrially supported software. Additionally, I19 also has another Dectris Pilatus 300K HPAD on a separate diffractometer setup (Instrument EH2).

Crystallographers may have software suites that they prefer to use and often want to process their data in software they are more familiar with (and know more nuances of) to enable the best possible data processing result. In the case of the North East X-ray Team who often have experimental time on beamline I19, being able to process data using the *APEX2*[197] (and now *APEX3*) suite of programs is preferred. However, at the time this software

did not support the *.cbf* frame format output by both detectors. In order to use the *APEX* suite of programs, a piece of software was written to convert the *.cbf* frames into *.sfrm* format (Bruker AXS Ltd standard frame format).

7.1 Data conversion

A program, *cbf_to_sfrm*, has been written in *Python2.7* to convert the *.cbf* frames output from the two diffractometers on beamline I19 into *.sfrm* format. As one file conversion does not depend on another, the process can be parallelised to speed up overall data conversion. Parallel computing allows these types of processes to be carried out simultaneously.[198] In this case, the number of frames to be converted is divided between separate processers on a computer or cluster. Running the conversions separately often ensures the overall conversion is completed faster than simply converting one frame after the other.

At Diamond Light Source, the conversion code can be run on a computer cluster – allowing the process to be divided amongst any number of processors (in practise 40 are used). On a conventional PC, this is also true, although conventional wisdom suggested that the maximum number of processes that can be run is no more than the number of available processors on the computer. Any more than this and the separate programs will split the percentage use of a processor (i.e. 2 programs running on one processor, each at 50% of the processing speed) and there will be no further increase in conversion speed. Two separate versions of the code exist – one for use at I19 and the other for use on a home PC – as the method for parallelisation is slightly different for each. The overall action of the code, however, is the same and will be detailed below with the individual modifications for each purpose explained thereafter.

Converting the data to *.sfrm* format produces very large data files and therefore the program also provides the option to compress data by adding up intensities for each pixel over the desired range of frames and then outputting an *.sfrm* file which contains the overall pixel intensities for that range. This is also a feature of the *DIALS* integration program,[199] which was developed for use on synchrotron beamlines. In practise (as shown in Section 7.1.5) the compression has some effect increasing the values of refinement statistics and decreasing the significance of the data – but has a large effect in decreasing the overall size of data that must be copied back from the beamline to process in the home laboratory of the user. The compression used is often a compromise between these two factors.

Data conversion works in two stages – firstly, finding the locations of all the frames to be converted and secondly, producing files with information from frames in *.cbf* format converted into *.sfrm* format. The original *.cbf* images are not overwritten by the program so the original data is not lost or altered in any way, indeed this is a requirement of the data architecture at Diamond Light Source.

7.1.1 Required input

In order to convert the diffraction frames, the user is required to supply the program with a series of inputs. These are:

- The location of one of the *.cbf* files to be converted. The code assumes that all the files for an experiment are in the same folder. Currently the program can only convert data from one diffraction experiment at a time.
- The prefix (info before the run and frame number) of the file e.g. *exptinfo* from a file called *exptinfo_01_00001.cbf.*
- The location of the default file (described in Section 7.1.2) which provides the program with additional information about the equipment not included in *.cbf* file.
- The location where *.sfrm* files should be written to.
- The number of runs to be converted. This assumes that the run numbering starts at 1.
- The number of processes or cores for parallelisation.
- The number of frames to be compressed together. If no compression is required, the value should be set to 1.

7.1.2 Defaults file

The *.sfrm* format requires a fixed set of information (or metadata) pertaining to the experiment to be included in the image header,[200] not all of which is included in the header of the original *.cbf* file. To remedy this, a default file containing all the additional metadata which must be included is created. For the purpose of transparency, it was chosen to input the information in the form of a text file instead of writing it directly into the code. This allows any users of the program to see what additional information is being added to the files (e.g. detector position corrections) that may affect the processing of the data and to easily change those values if they are aware of better choices for those parameters.

APEX2 will only process data from a specific list of detectors, for which it has additional correction information (e.g. phosphor thickness) hard-coded into the software. To bypass

this problem, the detector type of the converted frames is set to 'CMOS-PHOTON100' as it was the only detector type large enough to accommodate the image size of the large 2M detector (1400 x 1694 pixels). As a consequence, there are some pre-set values that need to be changed before the data can be integrated with *SAINT*. As Pilatus pixel hybrid detectors do not use phosphors, the phosphor efficiency of the detector must be set to 1.0 so the phosphor absorption correction is not applied.

The diffractometer goniometer geometry (kappa or Eulerian) must be specified in the default file to enable the program to convert the experimental angles correctly. Additional angle offsets from standard Eulerian geometry specific to the two diffractometer geometries are also recorded in the default file.

7.1.3 Splitting jobs

The first stage of the conversion is locating all the frames that are required to be converted. This section of the program uses the frame file location, number of runs and the frame file prefix (provided in the input file) to search for the number of frames with the given prefix in each run in the specified location. A list is created containing the locations of all the files in each experimental run and output as a text file to be used later in the conversion process. If compression of the frames is required, the list will only contain the name of the first frame in the list of frames to be compressed together; the program assumes that the frames will be numbered in a consecutive order and therefore additional file names are not required.

7.1.4 Diffraction image conversion

As integration software requires consecutive frames in order to process the data, a decision was made to convert the frames in a consecutive order. This is useful for large datasets, so it is possible to begin processing data before the conversion of the whole dataset is complete. The text file of frames to be converted that was created in the previous step lists the frames in order of increasing file number, allowing the frames to be converted in order.

To convert a diffraction frame into *.sfrm* format, the metadata from the *.cbf* frames must first be extracted. To accomplish this, a script written by Marcus Mueller for Dectris Ltd.,[196] '*pilatus_header.py'*, is used with slight changes to make it compatible with the *cbf_to_sfrm* program.

The next stage is to extract the image information, which is stored in binary code; *.cbf* image data is compressed to reduce the overall size of the diffraction frame files – the type of

compression scheme used for storing image data is specified within the frame header.[201] All of the frames created at I19 are stored with byte offset compression; instead of a list of the actual intensity values of each pixel, the image values are stored as the difference between a pixel and the previous one. A description of this encoding is included in Section 10.5. The extracted image information is stored in a matrix of a size specified by the header. If frames are to be compressed the program will iterate through the frames to be combined and add the intensity value of the pixel to the corresponding value in the matrix (so that all matrix items contain the sum of the intensity values of that pixel on each frame). If the number of frames in a run is not divisible by the desired compression, the last few frames that constitute the remainder will not be converted.

As previously stated, *APEX2* can only integrate data with a square image of a multiple of 512 pixels wide. This requires the image data to be padded to a multiple of 512 (1024 for 300K detector, 2048 for 2M). Figure 7.1 shows a converted image from the 2M detector as it appears in *APEX3* software; the black area to the top and left of the image is padding. Most Pilatus detectors are made up of an array of readout chips, therefore the black lines within the image are the areas between chips on the detector, where no data can be collected.

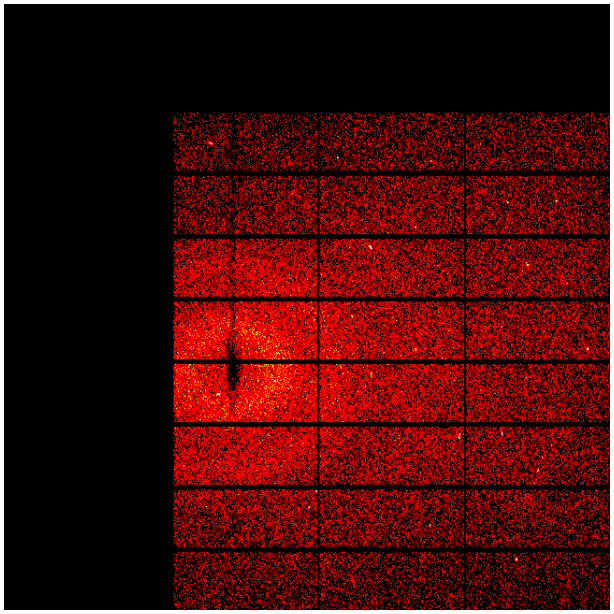


Figure 7.1 An image of one converted frame for instrument EH1 with Pilatus 2M detector. The black area at top and left of image is frame padding.

To account for the padding in the integration stage of the refinement and the spaces between the chips, active masks for each detector have been created. These files indicate to *SAINT* which areas of the frame to integrate. Figure 7.2 shows an image of the active mask

file created for the 2M detector. The areas in red are the only areas that will be integrated when this file is processed.

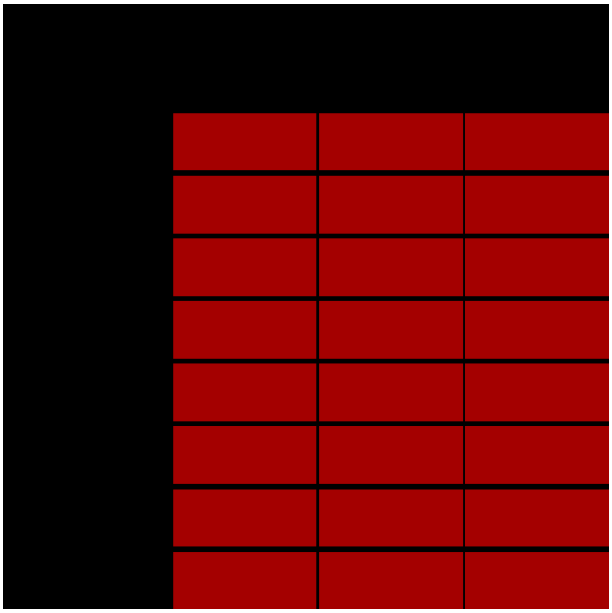


Figure 7.2 Active mask for instrument EH1 with Pilatus 2M detector. The dark area is area that will not be integrated. The grid motif due to the space between individual readout chips of the detector.

The final image matrix is then converted into the format required for *.sfrm* files (detailed in Section 10.6). All values below 0 within the image matrix are set to 0 to avoid using the underflow table within the *.sfrm* file – this will have no effect on integration as these values occur on areas of the detector between the chips and will not be used due to the application of the active mask.

Finally, the header information from the *.cbf* files is taken and converted into the form required for *.sfrm* format. One of the aspects of paramount importance within this conversion process is the transformation of the diffractometer angles. The zero positions of the rotation axes and the direction of the positive rotation of a goniometer setup vary dependant on the manufacturer and this must be taken into account when converting between *.cbf* and *.sfrm* formats.[202] The *.sfrm* format requires angles to be in Eulerian geometry (using ω, ϕ, χ, 2θ angles) therefore the diffraction angles for Instrument EH2 must undergo further conversion as the movement of the detector is described by a different set of angles known as kappa geometry (using ω, ϕ, κ, 2θ angles).

Instrument EH1 is in Euleurian geometry, however angle conversion is still required to ensure the axes of rotation fit the Bruker standard which is different to the standard definition of Eulerian geometry.[202] The conversion is thus:

$$\omega_{\mathrm{sfrm}} = -(\omega_{\mathrm{cbf}} + \omega_{\mathrm{offsett}}) \quad 7.1$$

$$\chi_{\mathrm{sfrm}} = -\chi_{\mathrm{cbf}} \quad 7.2$$

$$\phi_{sfrm} = \phi_{cbf} \quad 7.3$$

$$2\theta_{\mathrm{sfrm}} = -2\theta_{cbf} \quad 7.4$$

Here the subscript denotes whether the angle comes from a *.cbf* or *.sfrm* file and the offset is defined in the defaults file. The current omega offset used is 180°.

For instrument EH2, the following formulae are applied to convert from kappa to Euleurian geometry:

$$\omega_{sfrm} = (\omega_{cbf} + \delta) + \omega_{\mathrm{offsett}} + 180 \quad 7.5$$

$$\chi = -(2\sin^{-1}(\sin\left(\frac{\kappa}{2}\right)\sin(\kappa_{\mathrm{offsett}}))) \quad 7.6$$

$$\phi_{sfrm} = -(\phi_{cbf} + \delta + 90) \quad 7.7$$

$$2\theta_{sfrm} = 2\theta_{cbf} \quad 7.8$$

$$\text{with } \delta = \tan^{-1}(\tan\left(\frac{\kappa}{2}\right)\cos(\kappa_{\mathrm{offsett}})) \quad 7.9$$

The current omega offset is 90° and $\kappa_{\mathrm{offsett}}$ is 50°.

7.1.4.1 Individual PC

The original version of the code was created to be run on a PC. This has been parallelised using the multiprocessing function in Python and has been tested using the Scientific Linux 7.0 operating system.

7.1.4.2 Diamond

To run on the computer cluster at Diamond and decrease file conversion times to be in line with the average experimental time (15 minutes),[195] the code is split into two separate scripts. *split_jobs.py*, retrieves the names of files to be converted and outputs them to a text file. Because this text file is required for use in the next stage of the conversion (and is always given the same name) only one dataset can be converted at a time on the computer cluster or the program risks extracting data from the wrong experiment. The second step uses a bash script that must be submitted to the computer cluster which imports the relevant python modules needed before running the conversion (*cbf_to_sfrm_diamond.py*). To convert the files in consecutive order the conversion program iterates through the text file and takes every jth file, where j is total number of jobs, starting from the job number it is given (1:j). This code requires the input and default file.

Feedback on the program by users indicated that editing input scripts and running two separate commands could be confusing. Therefore, additional scripts were created for accessibility. *setupconversion.py* creates the bash and input files with correct file names and folder locations based on the specific architecture of folders at a synchrotron visit to beamline I19. This script only needs to be run once at the start of the conversion process. *allatonce.py* allows the whole conversion process to be run using four command line inputs (folder, prefix, output-folder, compression). It writes to the input file before running *splitjobs.py* and then submitting the jobs to the computer cluster.

7.1.5 Verification

To ensure the program converted frames correctly and there was no incorrect data manipulation, data for single crystals of L-Cysteine were collected on both diffractometers. As L-Cysteine is a chiral molecule, the confirmation of the correct absolute structure using the data converted by *cbf_to_sfrm* indicated that no incorrect data manipulation had occurred during the conversion process on either experimental setup. L-Cysteine has 4 polymorphs; the single crystal measured is polymorph I which crystallises in the orthorhombic $P2_12_12_1$ space group.

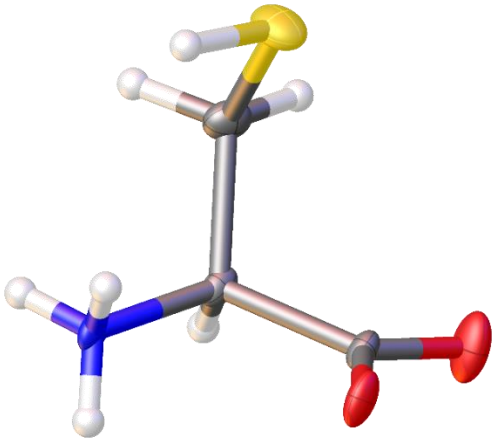


Figure 7.3 L-Cysteine as refined using cbf_to_sfrm.

7.1.6 Effect of compression

The effect of the frame compression was investigated on Instrument EH1 using a single crystal of L-Cysteine. The experiment took place at 100 K with λ = 0.68890 Å. Four experimental runs were collected, one phi and three omega scans (at the time, the standard North East X-ray Team collection). The frames were converted with compression levels of 1, 2, 4, 6 and 8 to give scan lengths of 0.1, 0.2, 0.4, 0.6 and 0.8 degrees respectively. These frames were then processed using the *SHELXL* suite of programs within the *APEX3* GUI. The data were integrated in *SAINT*, used in the LINUX command line, with the correct active

mask and the simple sum only algorithm and the phosphor correction turned off. The data were corrected for absorption and scaled in *SADABS* and the space group determined in *XPREP*. The structure was solved using *XT* and *XL* was used to refine the model within the *Olex2* GUI.

Compression	1	2	4	6	8
R_1 (%)	3.98	4.37	4.29	4.36	4.42
R_{int} (%)	5.68	6.11	6.46	7.10	6.94
wR_2 (%)	10.59	11.67	11.33	11.52	11.85
GooF	1.096	1.076	1.085	1.088	1.085
I/σ_I	25.3	23.1	22.1	20.5	19.4
Weighting parameters	a = 0.070 b = 0.142	a = 0.081 b = 0.159	a = 0.074 b = 0.176	a = 0.079 b = 0.110	a = 0.079 b = 0.039

Table 7.1 Table of refinement statistics from processing L-Cysteine with 1,2,4,6 and 8 levels of frame compression.

A series of refinement statistics are listed in Table 7.1. A generally increasing trend is seen for R_1 with increasing frame compression. This is also the case for R_{int} and wR_2.

Compression	1	2	4	6	8
a (Å)	5.4493(6)	5.4487(6)	5.4460(6)	5.4466(6)	5.4479(6)
b (Å)	8.1694(9)	8.1723(9)	8.1693(9)	8.1703(9)	8.1723(9)
c (Å)	12.0771(12)	12.0865(13)	12.0749(13)	12.0771(13)	12.0811(13)
V ($Å^3$)	537.64(10)	538.19(10)	537.21(10)	537.44(10)	537.87(10)

Table 7.2 Unit cell parameters for L-Cysteine with 1,2,4,6 and 8 levels of frame compression.

The unit cell parameters for each compression are reported in Table 7.2. All parameters are within statistical error of each other, except the value of c for a compression of 2 versus 1, 4 and 6. The refinement and unit cell parameters for compression level 2 break with the trend of the data. This may indicate a potential issue with the creation or processing of the data, the cause of which is currently unknown.

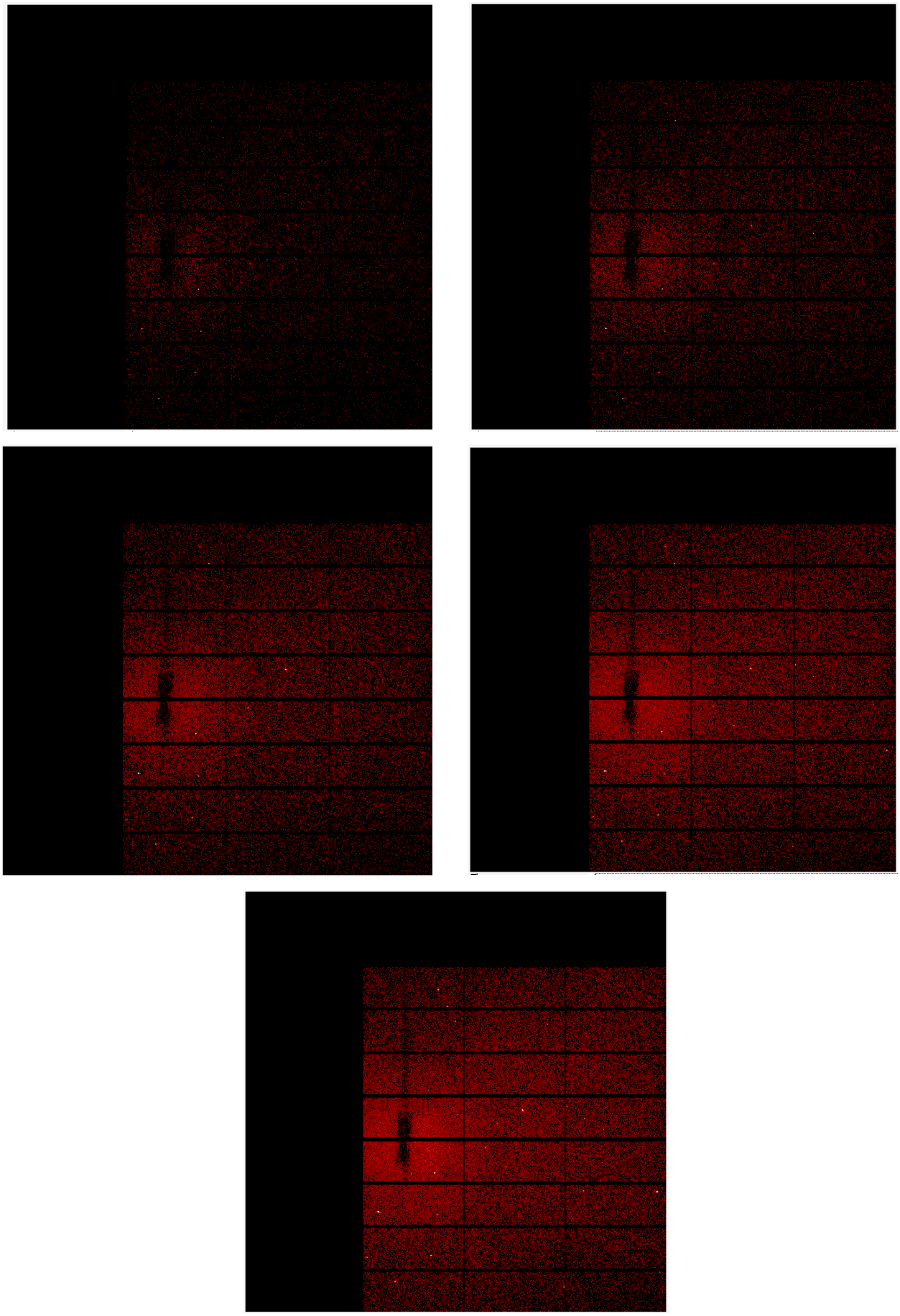


Figure 7.4 Image of the same frame of L-Cysteine with 1, 2, 4, 6 and 8 level of compression. The same intensity scale is used for all images.

The same frame at each compression level is displayed in Figure 7.4 with the same colour scheme for the display of the intensity values applied. Increasing the compression increases the background on each frame, as shown by the increasing brightness of the image from compression level 1 to 8. Diffraction from the crystals of L-Cysteine when no compression is applied is difficult to see due to the low background, but this frame is included for completeness. An increase in the background could be a reason for the decrease in significance, I/σ_I, seen in Table 7.1. Increasing background could cause the standard uncertainties for reflection measurements to increase, thus lowering the overall I/σ_I.

For this particular data collection, the crystal moved between experimental runs. This led to difficulty processing the data when no compression was applied (level 1). The first run had to be integrated with a different unit cell to the other three runs to ensure that the integration could locate the reflections in each run correctly. Integrating with one unit cell for all four runs resulted in very few reflections being located from run 1. This problem did not occur in later compressions, all from compression level 2 and above were integrated with one unit cell for all four runs.

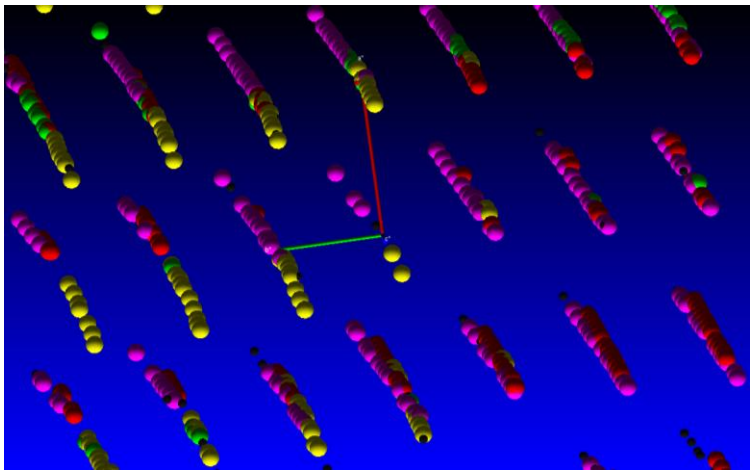


Figure 7.5 Reciprocal lattice of compression 1 – each colour represents data from a separate experimental run.

The reciprocal lattice when no compression was applied is displayed in Figure 7.5; reflections from each run are shown in a different colour. The reflections from different runs are not coincident, which suggests the crystal has moved. At a compression of 2 (Figure 7.6) the effect does not appear as pronounced.

One possible explanation for the improved alignment of the reciprocal lattice when the diffraction frames were compressed could be that the compression of the frames resulted in an increase in the ambiguity of the centroids of the reflections to such an extent that the

effect of the crystal moving was reduced. The precise location of the reflections (due to thin slicing of frames) and the low point spread function of the HPAD detector leads to reflections spread over only a few pixels. Therefore, a small displacement of the crystal can have a large effect on the ability of the integration program to locate reflections.

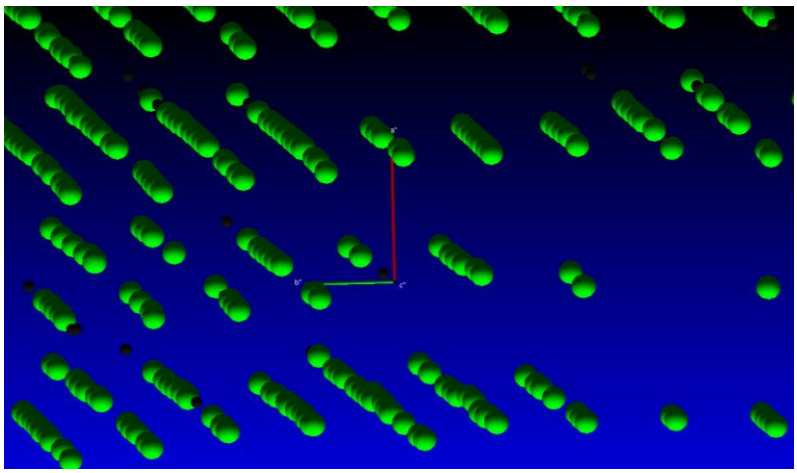


Figure 7.6 Reciprocal lattice of compression 2 in same direction as compression 1 - reflections from different runs are all same colour.

7.1.7 Limitations of cbf_to_sfrm and future work

As mentioned above, due to the way in which *cbf_to_sfrm* creates a list of file locations for the parallelisation of the conversion on the computer clusters at Diamond, it is only possible to run one conversion at a time. This could be prevented by creating a text file with a different name for each conversion being run, as well as creating individual different bash files for each conversion. This may be a useful modification for converting data at speed.

It may also be useful to a user to be able to convert specified runs. The program assumes that run number of the frames to be converted starts at 1. Selected non-consecutive runs (e.g. 2, 3 and 6) cannot be converted separately.

The *APEX2* integration software is only capable of integrating square diffraction frames. However, this is not the case for updated *APEX3* software, therefore, the size of each diffraction frame could be reduced by removing the image padding that is currently applied. This would require updating of the active masks and beam centre calculations.

cbf_to_sfrm could also benefit from further optimisation to improve the speed of data conversion. When the program is operated using the computer clusters at Diamond it keeps pace with a standard experimental collection, due to the number of jobs which can be submitted to the cluster, but on a conventional computer, which has access to many fewer processors, a conversion can take a lot longer. The program has not yet been fully optimised

in order to ensure the conversion process is as efficient as possible. Conversion speeds may also be improved by rewriting the program in a faster computer language than *Python2.7*.

Chapter 8. Diffraction data replication

Another avenue of investigation that was considered was a thorough analysis of the integration stage of the crystallographic experiment to identify any sources of systematic error and to examine the impact of those errors on the final multipole model. There are many potential sources of error that can arise during an experimental data collection, although it is the function of the data reduction and later processing software to attempt to correct for these errors. However, during the application of corrections or through the course of refinement of the data, systematic errors can occur within a dataset. A useful study would be to analyse the effectiveness of the integration stage alongside the effects of experimental noise or other experimental features in the data.

A methodical approach would involve using synthetic diffraction data, which would allow control of the level of systematic and random noise within any given diffraction image, as well as the form of the experimental background. This would enable the study of their effects on the outcome of data reduction. Synthetic diffraction data has already been put to use to probe macromolecular crystallography and different programs are available for the simulation of diffraction data with varying options, outputs and focuses.[203–207]

While there are several software packages for the simulation of single crystal diffraction data available, they were created for the macromolecular field.[204–206] Additionally, the source code of the existing programs are not readily available, limiting the inclusion of additional corrections or sources of error tailored to the expected experimental output from the diffraction of small molecule single crystals in order to fully probe the refinement process. Thus, the creation of an original piece of software was sought to allow full control over the precise format of the experimental noise and the background.

To develop and improve any findings, access would be required to the source code of the integration algorithms and any corrections that were applied by a particular software. These are not currently available as they are predominantly provided by instrument manufacturers. An initial test program, *Replica*, was created with a view to carry out initial testing without access to this source code with the hope of identifying any larger systematic

errors present. This work inadvertently led to a discovery of a facile route to create fraudulent diffraction data, highlighting the potential impact on the community if this were to take place.

The software described within this chapter is the result of a preliminary test case in order to investigate the difficulty of producing synthetic diffraction frames for small molecule crystal structures. The program is described with the view to demonstrate a simple mechanism for creating diffraction frames in reference to the production of fraudulent diffraction data, rather than how well it reproduces experimental results. In order to carry out the study as described in the introduction to this chapter, much work would be required to improve the synthetic frames created to give a more realistic output.

8.1 Creating synthetic diffraction data

Replica is written in *Python2.7* and requires information from an existing data collection, *.raw* and diffraction data *.sfrm* files, along with an input file (Section 8.1.3) detailing different experimental parameters to be applied in the frame creation. The frames are output in *.sfrm* format, as utilised by Bruker Axs Ltd. This was chosen as the *.sfrm* format is fixed and well documented and the format in which most of the diffractometers in the Newcastle University Crystallography Laboratory, at the time of writing, output their diffraction frames. It is also the format that can be used with *APEX2* to integrate the data.

Each image within a diffraction experiment is the result of the measurement of the intensity of a diffracted X-ray beam for particular crystal and detector positions. In order to recreate this, reflections which have satisfied the Bragg equation at that particular combination of crystal and detector angles must be known. This would ordinarily require the calculation of scattering vectors for each of the reflections within a dataset, at different setting angles of the goniometer, to provide the position where each reflection would be measured. This is a labour-intensive calculation requiring the knowledge of the orientation matrix of the crystal. However, to recreate an experiment which has already taken place, this is not necessary as the information can be extracted from previous files.

The *.raw* files, created after integration by *APEX2*, contain a wealth of information about each measured reflection – including the position at which the reflection was measured on the detector. This information can thus be used to circumvent the need for calculating the positions of reflections. Therefore, only the spreading of the reflection intensity over space

needs to be calculated for the replication of diffraction frames. This enables experimental datasets collected on laboratory instruments to be replicated with the potential to vary different parameters to probe the integration stage of structure solution.

Making diffraction frames can thus be divided into two steps. For every reflection, the amount of the total intensity of the reflection that will appear on each frame is calculated. This information is then transformed into a 2D representation of a diffraction pattern to produce the diffraction frames.

8.1.1 Intensity allocation

As eluded to previously, *Replica* uses a combined approach to generate synthetic data; the header, which contains information about the experiment, is taken from the existing diffraction frames and combined with the position of reflections and intensity values derived from the refined structure from the same experiment to produce the diffraction frames. As the positions of reflections are known from existing experimental data, the only calculation required is assessing what portion of the total reflection intensity lies at that point on each individual frame. This is presently achieved using simple normal distribution calculations.

The reproduction begins by obtaining the requisite values from the header of an existing *.sfrm* file, including the number of bytes per pixel used to store the diffraction intensity values. The overall image dimensions are required to evaluate whether the pixel coordinates in the *.raw* file have been scaled. In addition, the beam centre and distance of the detector from the sample are used to calculate any corrections to pre-apply. The information for each reflection is extracted from the *.raw* file and stored as a list.

The program then calculates which reflections will appear on each diffraction frame. Within the *.raw* file, the position of the centre of the reflection is described by 3 coordinates (*x, y, z*). The *x* and *y* coordinates describe the location of the reflection on the 2D diffraction image, while the *z* coordinate gives the frame number on which the centre of the reflection is located. Intensity relating to an individual reflection usually appears on more than one diffraction frame; to account for this, the total intensity value of the reflection must be divided between consecutive frames.

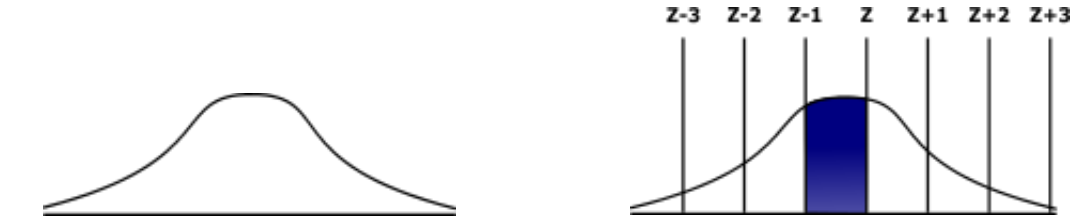
To spread the intensity of a reflection over a number of frames, the profile or shape of the reflection in the *z* direction is modelled as a perfect Gaussian distribution, described by Equation 8.1, the shape of which can be seen in Figure 8.1a.

$$f(z|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(z-\mu)^2}{2\sigma^2}} \quad 8.1$$

The value of the Gaussian at a position, *z*, depends upon the spread of the function – controlled by the standard deviation, σ, and the difference of z and the mean value, μ. In *Replica*, σ is defined by the user, σ_z, and z coordinate of the reflection centroid, z_c, is used as μ (Equation 8.2).

$$f(z|\mu,\sigma) = \frac{1}{\sigma_z\sqrt{2\pi}} e^{-\frac{(z-z_c)^2}{2\sigma_z^2}} \quad 8.2$$

The percentage of the Gaussian between each consecutive frame is assumed to be the same as the percentage of the intensity value that appears on the later frame. In Figure 8.1b, the blue shading between frames *z* - 1 and *z* indicates the percentage of the total intensity that will be attributed to the reflection on frame *z*.



a) Gaussian shape of replicated Z direction profile. *b) Total intensity of reflection on frame Z in blue.*

Figure 8.1 Calculation of the total intensity of a reflection per frame.

To calculate the value of the total intensity on each frame the *s* score (usually called the z score in statistics), the number of standard deviations the frame number is from the z_c, is calculated using Equation 8.3, where z is the frame number.

$$s = \frac{z - z_c}{\sigma_z} \quad 8.3$$

The s score can be translated into a p value using the conversion factor (Equation 8.4) to gain the percentage area under the Gaussian distribution up to that value of *s*.

$$p = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{s}{\sqrt{2}}\right)\right) \quad 8.4$$

$$\mathrm{erf}\left(\frac{s}{\sqrt{2}}\right) = \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{s}{\sqrt{2}}\right)^2} \quad 8.5$$

The p value for the earlier frame (z – 1) is then taken away from the latter frame (z) leaving the difference – the percentage of the distribution between the two frames. The total intensity for the reflection on that frame, I_z, is then calculated from Equation 8.6.

$$I_z = I_{tot} \times p_{z-1:z} \quad 8.6$$

The program first iterates for increasing values of z from the central pixel of the unit cell calculating values for the percentage until I_z rounds to 0. The calculation is then performed with decreasing values of *z* until the same condition is met; I_z, along with the coordinates of the x and y position of the centre of the reflection, is stored in a list for the corresponding frame. After the intensity spread is calculated for each reflection in the *.raw* file, there will be a list for every frame containing the position and intensity of every reflection present on that particular frame.

8.1.2 Frame creation

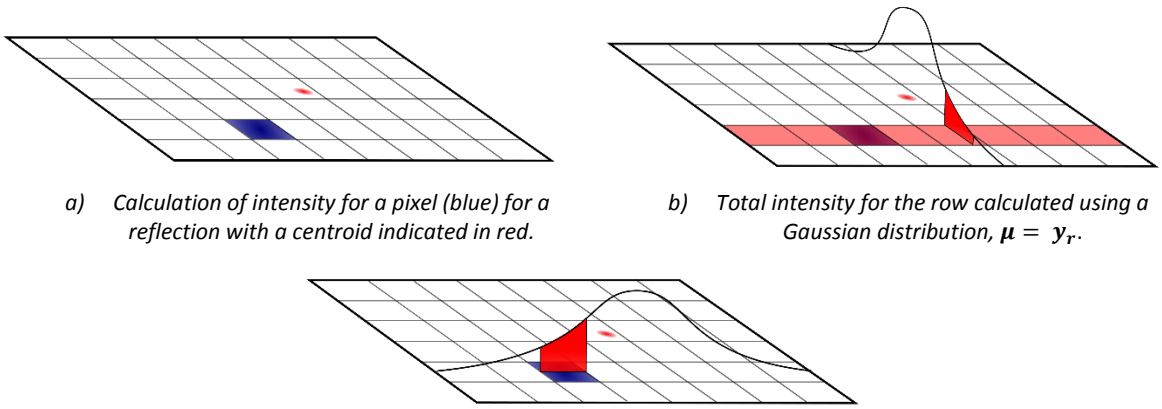
The diffraction image data is then constructed for each frame. Within a frame the intensity relating to a reflection is recorded on more than one pixel due to physical characteristics of the single crystal, such as its size and mosaicity. Therefore, the total intensity of the reflection on that frame needs to be spread in 2D across a number of pixels. A similar process to the one outlined above is employed to calculate a given intensity of a reflection for each individual pixel on an image using σ_{xy} as the standard deviation of the reflection profile on the frame in place of σ_z.

The maximum spread of the reflection is calculated using Equation 8.7, where I_{tot} is the total intensity of the reflection on the frame in question and σ is the mosaicity in the x and y directions. In the current iteration of the code the y spread is assumed to be identical to the x spread. The integer of max(x) is the number of rows and columns to iterate through when spreading the spot on the frame. Equation 8.7 gives the number of pixels from the centre of the reflection that will give an integer value of 1 or above.

$$\max(x) = -2\sigma_{xy}^2 \times \log\frac{\sigma_{xy}\sqrt{2\pi}}{I_{tot}} \qquad 8.7$$

The process of spreading the reflection intensity in 2D is illustrated in Figure 8.2. To calculate the value of the intensity for a pixel (Figure 8.2a), first the percentage of the intensity of the reflection for row the pixel is on is calculated (Figure 8.2b) using the method described above (Equations 8.3, 8.4 and 8.6). The 2D reflection profile is modelled as 2 Gaussian distributions, where the value for μ is either the x and y coordinate of the centre of the reflection. Whole number coordinates within the image are assumed to be at the centre of each pixel, so the percentage of the distribution for a row is calculated between $\pm$ 0.5 of the y coordinate.

This procedure is then repeated to find the intensity for a particular pixel by determining the percentage of the Gaussian between x $\pm$ 0.5, then multiplying it by the total intensity calculated for the row (Figure 8.2c). Once this has been completed for every reflection present on the frame, any required experimental background or noise is added to the image matrix. To create a background for the synthetic data, a very basic approach was taken. A noise value is added to every pixel in the image after all the intensity values are calculated. The value of the noise is a random number chosen for each pixel from a range that is defined in the program.



a) *Calculation of intensity for a pixel (blue) for a reflection with a centroid indicated in red.*

b) *Total intensity for the row calculated using a Gaussian distribution,* $\boldsymbol{\mu} = \boldsymbol{y_r}$.

c) *Total intensity for the pixel then calculated using another Gaussian,* $\boldsymbol{\mu} = \boldsymbol{x_r}$.

Figure 8.2 Spreading the total intensity of a reflection on a frame.

The frame is then written out in the required format; the header from the original experimental frame is used, but the image data is replaced with the calculated values. The total intensity of the frame, found in the header information, is also replaced with a value calculated from the sum of the new image data. These frames can then be processed in the same manner as the original data.

8.1.3 Input file

The input file required for *Replica* contains the location of the *.raw* file from the experimental data reduction along with the file name of an existing *.sfrm* file and the folder the frames are located in.

Within the file, the desired standard deviation of the reaction profile in both the *z*, σ_z (between the frames), and *x* and *y* direction, σ_{xy} (within the frame) must be specified. In the *z* direction this affects how many frames the reflection appears on. For both deviations the specified value is applied to all reflections regardless of their position on the detector. As it was a preliminary test program, this approximation was used to simplify calculations within

Replica. The value of the total intensity for each reflection can also be specified as the raw peak sum, integrated peak sum or a constant value for all reflections.

Though not included in the input file, *Replica* also requires a string containing the desired fixed section of the output file name. *APEX2* expects files to be named in the following format: *fixed-string_run-number_frame-number.sfrm* e.g. test1_01_0001.sfrm, with 2 digits for the run number and 4 for the number of the diffraction frame within that particular run. The fixed string acts as a unique identifier for the frames and often contains information about the experiment.

8.2 An example of replica usage

An example of the output from *Replica* is described below. Experimental data of a single crystal of a sulphonium ylid, $C_{11}H_{10}O_2S$, was measured on the XIPHOS2 diffractometer[208] (equipped with a CCD) at Newcastle University. The measurement took place at 23°C, using a Ag source, λ = 0.560870 Å. The sulphonium ylid crystallises in monoclinic $P2_12_12_1$ space group. This crystal has a well characterised structure and is used to calibrate the machine.

The synthetic data were created using the *.raw* file produced when data collected from this crystal were integrated, with σ_z = 0.6 and σ_{xy} = 1.03 (based on values from the reduction of the real data). The positions of the reflections were taken from the observed values in the *raw* file and the multiplicative correction factor listed was applied to the intensities. Analysis of the background files for $C_{11}H_{10}O_2S$ indicated that the average background value was between 1 and 4 counts (not including the baseline offset value of the header file – which in this case was 32 counts). The background added to each pixel was therefore a randomly generated integer value between 0 and 4, plus an additional 32 counts for the baseline offset.

From a visual inspection of both the experimental and replicated data it is possible to see the presence of additional reflection spots in the replicated frames that are absent in the experimentally collected frames. For example, Figure 8.3 shows an experimental diffraction frame while Figure 8.4 is an image of the same frame created using *Replica* (highlighted extra reflections are shown in Figure 8.5). The extra reflections in the replicated data appear to be present in the next experimental data frame (Figure 8.6). This could suggest that a smaller value is required for σ_z, as the reflections are spread over too many frames when compared to the experimental data.

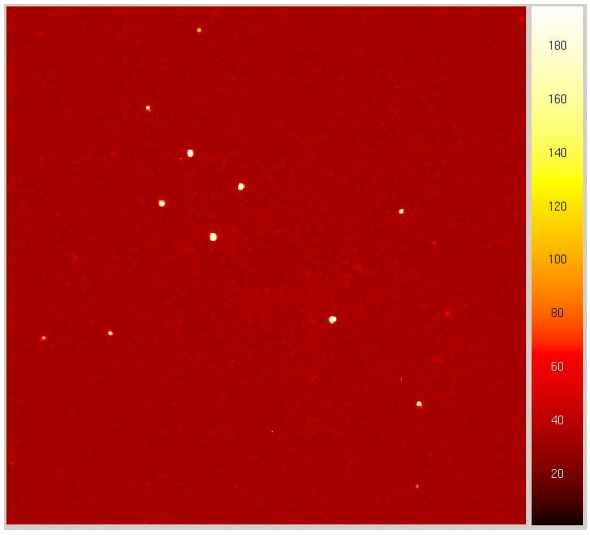


Figure 8.3 Frame 01_0414 from experimental data.

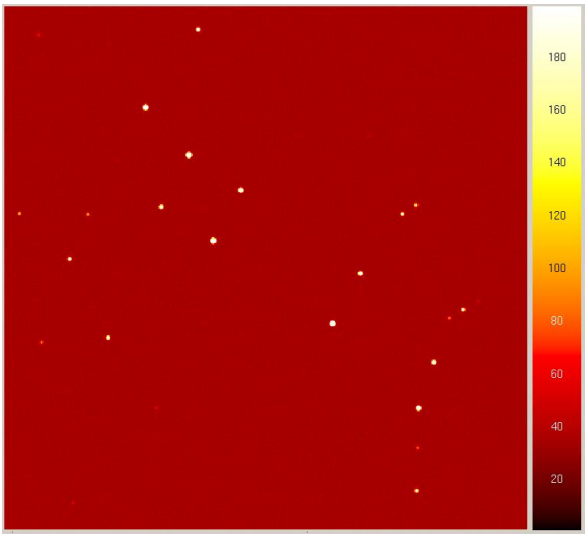


Figure 8.4 Frame 01_0414 from replicated data.

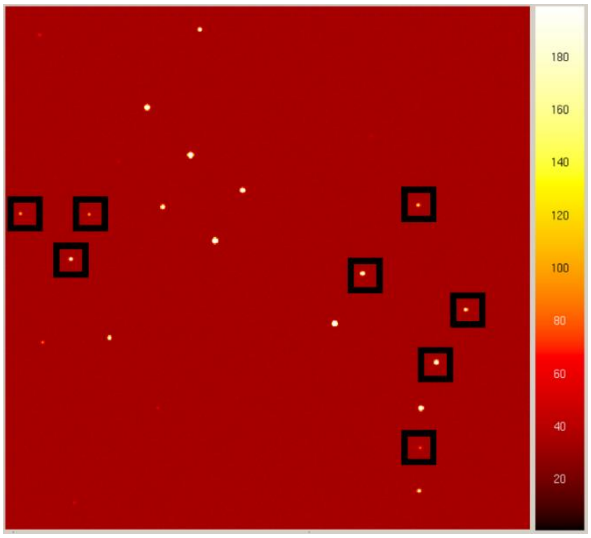


Figure 8.5 Frame 01_0414 from replicated data with additional reflections identified in black squares.

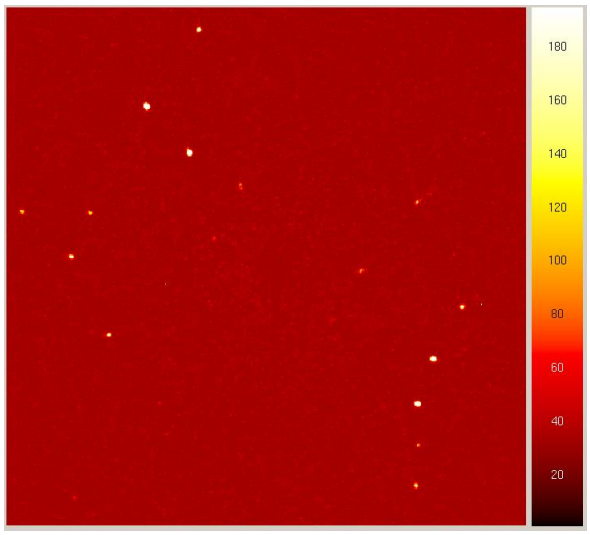


Figure 8.6 Frame 01_415 from experimental data.

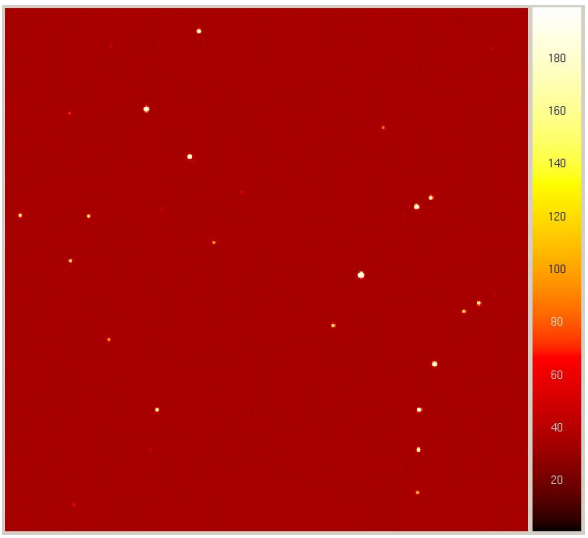


Figure 8.7 Frame 01_415 from replica data.

In general, reflections in the replicated data set appear on more frames than the experimental data. This could also suggest that the approximation of the same σ_z for all

reflections is too basic. The number of frames a reflection appears on in experimental data is not the same for all reflections and varies with reflection position. Reflections on the outside of the diffraction image would appear on fewer frames as they pass through the Ewald sphere more quickly. Other reasons could include a poor characterisation of the spreading of intensity between frames in the *Replica* program, or that the reflection centre calculated in the reduction program is not accurate. A better match is seen between Figure 8.6 and Figure 8.7, although some reflections are noticeably more intense in the replicated data.

Additionally, the frame background appears to be much noisier in the experimental data than the replicated data. Using the average background files, produced during the integration of real data, as a method for estimating the frame background gives a range that is too small and should be increased to give a more realistic representation of diffraction images. The spot shape also seems to be too perfect in the replicated data, on account of using an ideal Gaussian distribution to spread a reflections total intensity on a frame.

The previous data were from a non-standard collection, therefore a different calibration dataset collected with the same crystal was used to compare the output of refinements with real and replicated frames. Diffraction frames were created using *Replica* using the integrated intensity, as calculated in the *.raw* file, and reflection positions taken from either the observed or the predicted reflection positions. Frames were integrated in *APEX3* using the following options for the *SAINT* package: best-plane background and simple summation only integration algorithm to 0.7 Å. The simple summation algorithm was chosen to remove any differences resulting from the calculation of reflection profiles in the measurement of intensities. Data were scaled using *SADABS* with a default error model applied and solved and refined in *Olex2* using *XT* and *XL*.

	Real	**Replica (Obs)**	**Replica (Pred)**
a (Å)	5.9580(2)	5.9580(4)	5.9590(2)
b (Å)	9.0314(3)	9.0312(6)	9.0313(3)
c (Å)	18.3752(6)	18.3739(11)	18.3743(7)
V (Å^3)	988.75(6)	988.66(11)	988.86(6)

Table 8.1 Unit cell parameters from real and replicated model of a sulphonium ylid.

In Table 8.1, the unit cell parameters calculated from the real and replicated data sets are given. There are no statistical differences between any of the reported values, indicating the choice of reflection positions did not affect the calculation of the unit cell parameters.

A summary of the refinement statistics from the three models is presented in Table 8.2. The statistics look approximately comparable, although there is a decrease in the significance of the data created with *Replica*. The residual electron density present in the structure is comparable to the real data which may indicate that the global features of the model can be recreated using Replica.

	Real	Replica (Obs)	Replica (Pred)
R_1 (%)	4.15	3.87	3.86
wR_2 (%)	8.46	7.72	7.79
R_{int} (%)	7.15	7.89	7.79
R_σ (%)	6.13	8.81	8.47
Mean I/σ_I	16.3179	11.3486	11.8046
wGooF	0.977	0.883	0.881
Extinction correction	0.024(5)	0.031(5)	0.030(5)
Residual e peak ($eÅ^{-3}$)	0.2	0.2	0.2
Residual e hole ($eÅ^{-3}$)	-0.3	-0.2	-0.2
Weighting scheme	a = 0.036, b = 0.0	a = 0.031, b = 0.0	a =0.032, b = 0.0

Table 8.2 Summary of refinement statistics from real and replicated models of a sulphonium ylid.

8.2.1 Improvements

Replica was a preliminary piece of software designed to test the applicability of using the *.raw* file as a method to create replicated diffraction images. If it were to be further developed, there are many changes that would need to be applied to create more realistic diffraction images and produce a better match between experimental and replicated data.

While this initial test showed that it is feasible to produce diffraction images using the *.raw* file, the simple model used to spread the total reflection intensities would need to be refined. The ability to apply skewed Gaussian functions and a bivariate point spread function would be useful. The spread of the reflections due to the Lorentzian effect would also need to be taken into account so that reflections at higher resolution appeared on fewer consecutive frames. This could be introduced as a factor applied to the standard deviation of each Gaussian based on the Lorentz correction.

A more complex background would need to be added, perhaps with a constraint based on the surrounding pixels, so there was a smoother variation in the background across the frame. To investigate the effect of systematic and random error in the integration stage functions would need to be added to model these errors in different forms.

8.3 Fraudulent diffraction data

In 2010, an editorial in the journal Acta Crystallographica Section E notified members of the scientific community of a prolific case of fraud.[209] 70 papers were retracted from 2 authors

after it was found that crystal structures had been falsified.[210,211] In many cases, fraudulent transition metal or lanthanoid complexes were being created through the modification of a crystal structure from experimentally measured data by manually changing the unit cell parameters, as well as the heavy element in the complex in the model, whilst using the same structure factor file as the original experiment. In some lanthanoid structures there were cases where the ligands bound to the metal also had atoms changed. These structures were then submitted to the journal as if they had been measured experimentally. This 'atom substitution' also occurred for a number of organic structures. The fraud came to light by chance during the testing process of new software designed to check structural information for IUCr journals (of which Acta Crystallographica Section E is part of). Unexplained structural errors in the crystal refinement due to the metal substitutions lead to the discovery that the structure factor files for the structures were identical – thus several individual structures had been created from one set of data.

CheckCIF is the IUCr's crystal structure validation service.[212] All crystal structures submitted to IUCr journals as part of publications are required to be reviewed using the online checking system.[213] They are required to be in CIF format, the standard file for completed crystallographic refinements.[214] *CheckCIF* creates a report: a list of ALERTS for the structure of different levels of significance, based on how serious the problem is deemed to be. These range from missing experimental information in the CIF, to alerts which suggest structure may be incorrectly modelled or that the data used are likely of poor quality. This software is constantly being updated to include new checks.[215,216] Very serious alerts in a *CheckCIF* report are required to be explained within the CIF, as to why they have occurred, in order for them to be accepted for publication. *CheckCIF* reports can alert journal reviewers of issues with crystal structures that might need further assessment before publication, including fraudulent data.

After the 2010 case of fraudulent structures, a further 114 other publications were retracted by Acta Crystallographica Section E between 2010 and 2012 after using the new structural checks, all due to *"problems with data sets or incorrect atom assignments".*[217–221] These publications all contained mistakes in the treatment of the data that were said to be most likely accidental.[222]

Fraudulent data have been shown to be present across the wider scientific community, with many different fields looking into ways to prevent such data from being published. A check

for plagiarism is common for most academic journals. Many fields are also considering requiring researchers to make their raw data available,[223] although whether that raw data can be used to repeat the published experiment is not always enforced.[224] Allowing other researchers to reassess published raw data can allow that data to be reanalysed for different purposes other than that intended by the original researchers, to be used as part of a wider study to identify trends in data, or to help identify mistakes made by their peers. This process is already taking place using structure factor files processed from raw data, especially with macromolecular data.[225] Joosten *et al.* report that reprocessing existing macromolecular data with new techniques can give improved refinements, and noted that access to raw diffraction frames could improve future processing further as structure factor data has already been processed using *"best knowledge at time"* and new techniques could improve this processing.[226] This sentiment is echoed by Terwilliger *et al.*[227] who argue for deposition of raw data to aid continuous reprocessing of protein databank data to improve structure models. Deposition of raw data also has the added benefit of allowing members of the field to check for fraud or deliberate misinterpretation of any data that is published.

The Diffraction Data Deposition Working Group (DDDWG) was set up by the IUCr to "*address the growing calls within the crystallographic community for the deposition of primary diffraction images".*[228] If crystallographers were required to make their raw diffraction data available alongside completed structures, this would quickly add up to many petabytes of data, so the requirements need to be properly assessed – i.e. how will the data be stored, what format would the data be required to be in (different diffractometers and manufacturers output their diffraction data in different 'standard' formats), how would other crystallographers gain access to the raw data etc. alongside any metadata that would be required for reprocessing.[229,230] The initiative is ongoing, with many test cases of raw diffraction data being made available on the internet.[230–232] Alongside the desire for more open practises and the opportunity to reassess data, another advantage of the deposition of raw data as a method to prevent fraud was highlighted in a recent paper by Kroon-Batenburg and Helliwell: *"There is a further reason that is proposed for the utility of preserving raw diffraction data, namely the prevention of scientific fraud. The raw data would present a much greater hurdle against fabrication. The crystallographic community is somewhat divided on the effectiveness of this, however, in that it may ultimately prove achievable to fabricate raw diffraction data too."*[230]

As hypothesised by Kroon-Batenburg and Helliwell, the hurdle of requiring raw data is not one that is insurmountable. When using *Replica*, it became apparent that it would be possible to use the program to create raw diffraction frames for a non-existent structure with a few minor modifications to the code. Instead of using the intensities from real data found within the *.raw* file, calculated structure factors from a falsified structure could be used. These fraudulent diffraction frames could then be processed as normal, returning a structure that seemingly looks legitimate, with the raw data to apparently back it up. Such a case was highlighted at the DDDWG Satellite Meeting in Rovinj, Croatia at ECM29 in 2015.[233]

8.3.1 Fabricated data

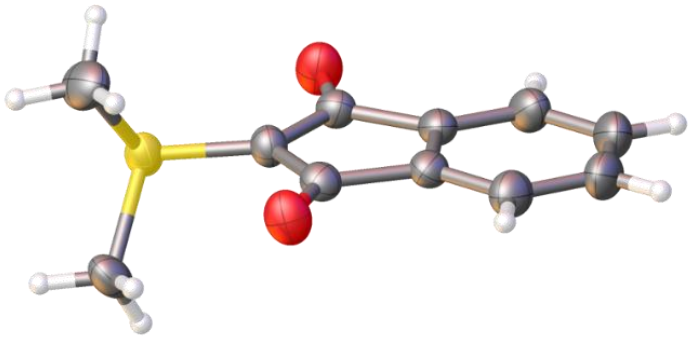


Figure 8.8 Structural model of sulfonium ylid (from a real diffraction experiment).

The second measurement of the sulfonium ylid, as used in Section 8.2, was utilised in this experiment. The data were processed to give a final structural model as indicated in Figure 8.8. Within *Olex2*, the final structural model was then modified manually, by deleting atoms and manually adjusting their positions to create a model of a fraudulent structure. The modified structure is displayed in Figure 8.9 with the following changes made:

- the sulphur atom has been changed to a selenium atom (and the selenium atom moved to make the C-Se bond longer),
- half of the 5 membered ring has been deleted,
- two carbon atoms within the benzene ring were changed to be a nitrogen and an oxygen respectively,
- additional experimental noise was included through the addition of partially positively and negatively occupied hydrogen atoms at random places throughout the asymmetric unit.

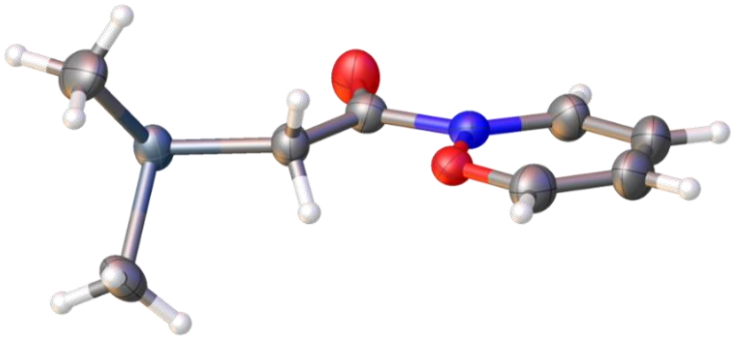


Figure 8.9 Proposed fraudulent molecule.

The *CheckCIF* report from the real sulphonium ylid data (Figure 8.10) shows only one error: a G level alert indicating that some reflections that may be expected to be present in this data are not there. However, this is a minor alert which would not be required to be explained when publishing this structure.

```
Alert level G
PLAT912 ALERT 4 G Missing # of FCF Reflections Above STh/L=  0.600        111 Note

   0 ALERT level A = Most likely a serious problem - resolve or explain
   0 ALERT level B = A potentially serious problem, consider carefully
   0 ALERT level C = Check. Ensure it is not caused by an omission or oversight
   1 ALERT level G = General information/check it is not something unexpected

   0 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
   0 ALERT type 2 Indicator that the structure model may be wrong or deficient
   0 ALERT type 3 Indicator that the structure quality may be low
   1 ALERT type 4 Improvement, methodology, query or suggestion
   0 ALERT type 5 Informative message, check
```

Figure 8.10 Section of CheckCIF Report from sulphonium ylid (real) created in August 2015.

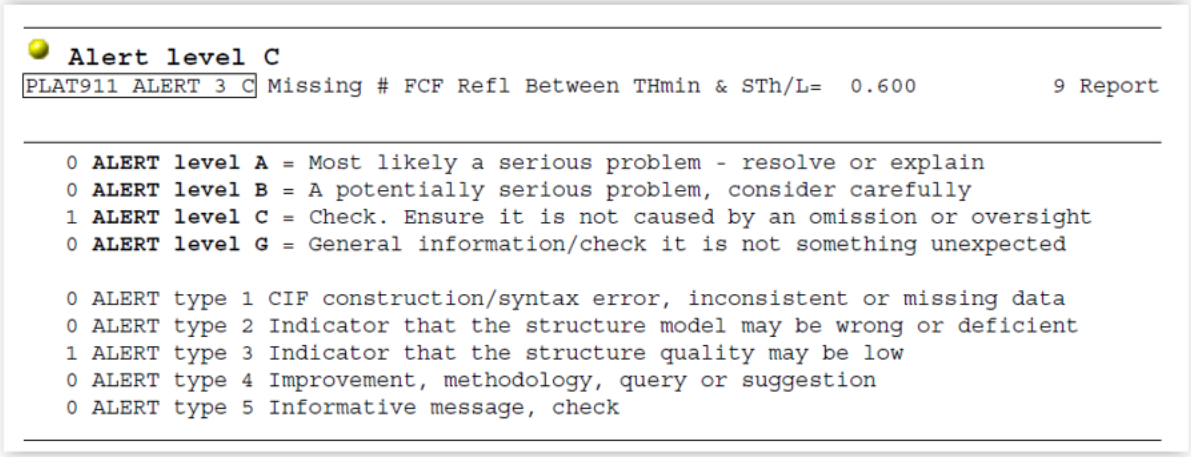
Compiling a CIF using the real sulphonium data with the fraudulent model produces a very different response. The *CheckCIF* report (Figure 8.11) shows many serious errors, including ones which suggest that the structural model is incorrect. This would signal to a reviewer that there is a fundamental problem with the structural model and prompt further investigation.

```
  32 ALERT level A = Most likely a serious problem - resolve or explain
  22 ALERT level B = A potentially serious problem, consider carefully
  72 ALERT level C = Check. Ensure it is not caused by an omission or oversight
   5 ALERT level G = General information/check it is not something unexpected

  18 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
  60 ALERT type 2 Indicator that the structure model may be wrong or deficient
  12 ALERT type 3 Indicator that the structure quality may be low
  41 ALERT type 4 Improvement, methodology, query or suggestion
   0 ALERT type 5 Informative message, check
```

Figure 8.11 Section of CheckCIF report from sulphonium ylid data, when structural model changed to proposed fraudulent model in Olex2, created in August 2015.

Using the calculated structure factors for the fraudulent model instead of total intensities from the real data allows us to create fraudulent diffraction images. These diffraction images were then processed using the same method as for the real data and returned a structure almost identical to the fraudulent model. When this data was refined and a *CheckCIF* report created, only one alert was present (Figure 8.12) for missing reflections – most likely a product of some weaker reflections being removed when recreating the code (as described in Section 8.1). This single error may not be enough to alert a reviewer that the structure has been falsified (although in this case an unreasonable structure might).



Alert level C
PLAT911_ALERT_3_C Missing # FCF Refl Between THmin & STh/L= 0.600 9 Report

0 **ALERT level A** = Most likely a serious problem - resolve or explain
0 **ALERT level B** = A potentially serious problem, consider carefully
1 **ALERT level C** = Check. Ensure it is not caused by an omission or oversight
0 **ALERT level G** = General information/check it is not something unexpected

0 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
0 ALERT type 2 Indicator that the structure model may be wrong or deficient
1 ALERT type 3 Indicator that the structure quality may be low
0 ALERT type 4 Improvement, methodology, query or suggestion
0 ALERT type 5 Informative message, check

Figure 8.12 CheckCIF report from fraudulent synthetic data, created in August 2015.

Select refinement statistics from the three structures are reported in Table 8.3. Using a fake model with the original sulphonium data resulted in an extremely high R_1 value, as expected. The data processed using *Replica*, however, had a low value for R_1, as it was created from a structural model which contained very little additional noise.

	Real	Fake model	Replica
R_1 (%)	3.73	38.01	1.11
GooF	1.047	4.502	1.044

Table 8.3 Refinement statistics for real, fake model and fake replicated data.

This fake model was created using an unsophisticated piece of software in an inelegant manner to give a deliberately absurd structure to showcase an extreme case of creating fraudulent diffraction data. Consequently, the unit cell from the falsified data was almost identical to the data from which it was created. A more sophisticated software would be required to calculate different positions for the reflections based on a new unit cell, but again that barrier is not insurmountable, especially when considering the synthetic software already available. It is also worth mentioning that there may be many unpublished structures at the disposal of a crystallographer which could be used as a starting point.

Another potential misuse of such software could be to tidy up noisy diffraction data, improving data quality indicators such as the R factor. Remaking a structure with an improved background could turn unpublishable data into a publishable structure determination. To draw an example from another field, biology journals have had to deal with the problem of photographs of experimental results being modified either to deliberately mislead or 'neaten up' data. In an effort to guard against image manipulation, the Journal of Cell Biology now sets out guidelines for what is deemed to be *"unacceptable image manipulation".* Rossner and Yamada describe the guidelines in their publication,[234] including examples that have ranged from the deliberately misleading – adding or removing features from images – or more subtle modifications – changing brightness or contrast on images or cleaning up the background of images – all of which can hide features of the data or alter conclusions. In the case of the Journal of Cell Biology, when they began screening images for modification it became apparent around 1% of papers accepted by the journal had images that had been modified in such a way that would alter conclusions of the paper.[235] With access to digital files, it is now more possible for editors (and reviewers) to check for doctored images.[236]

Within crystallography, even if access to raw data is provided, currently the only way to verify where the data originate from (e.g. machine, date measured) is by checking what is written in the diffraction frames themselves. This information is usually in a text format which could be modified leaving little trace. Any method to secure diffraction frames would need to include information from both the header and the image to ensure new image data was not simply pasted over the top of existing images. While raw diffraction data could be encrypted, this would slow down data processing (as it would need to be decrypted to be processed). This also moves away from the more open-source open-access model that science seems to be adopting.

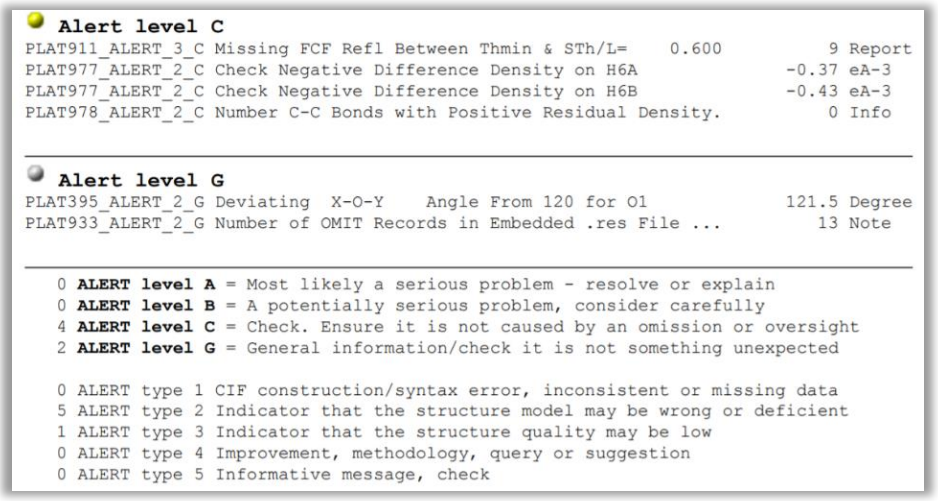
One more reasonable option could be a cryptographic hash function[237] which has many applications in information technology such as secure password storage, creation of digital signatures and integrity verification. A hash function is an algorithm which converts a piece of data of any size (the message/key) into a format with a fixed size. The output of the hash function is known as a hash. A cryptographic hash function is an algorithm with extra properties which make it extremely hard to take the output hash and regain the input message. This means that it is nearly impossible to determine the starting message value. To

ensure that the header and frame match (and a fraudulent image has not been used in conjunction with a header from a real experiment – as done in this example) the message to be hashed would need to contain information from both the image and the header e.g. intensities from certain pixels and then several choice parameters from the header.

This could work in a similar way to a digital signature, a method used in cryptography to detect if the contents of a document have been tampered with. The hash of the original document, created by the sender of the document, is compared to the hash of the document when it has been received by another party. If these are not the same, then the document has likely been tampered with in transit from the sender to the receiver. The hash included upon generation of a frame could be compared to a hash of the frame in the process of deposition into a databank to check for tampering. This would be secure provided the hash function was not common knowledge.

A similar function, called a checksum, already exists in *.sfrm* files and is included in the header. In this instance, the checksum is created by the summation of the intensity across the whole image. This value can easily be changed to match the data from a new image (as is performed by *Replica*).

8.3.2 Outcomes



Alert level C

```
PLAT911_ALERT_3_C Missing FCF Refl Between Thmin & STh/L=    0.600          9 Report
PLAT977_ALERT_2_C Check Negative Difference Density on H6A              -0.37 eA-3
PLAT977_ALERT_2_C Check Negative Difference Density on H6B              -0.43 eA-3
PLAT978_ALERT_2_C Number C-C Bonds with Positive Residual Density.          0 Info
```

Alert level G

```
PLAT395_ALERT_2_G Deviating  X-O-Y    Angle From 120 for O1             121.5 Degree
PLAT933_ALERT_2_G Number of OMIT Records in Embedded .res File ...         13 Note
```

```
   0 ALERT level A = Most likely a serious problem - resolve or explain
   0 ALERT level B = A potentially serious problem, consider carefully
   4 ALERT level C = Check. Ensure it is not caused by an omission or oversight
   2 ALERT level G = General information/check it is not something unexpected

   0 ALERT type 1 CIF construction/syntax error, inconsistent or missing data
   5 ALERT type 2 Indicator that the structure model may be wrong or deficient
   1 ALERT type 3 Indicator that the structure quality may be low
   0 ALERT type 4 Improvement, methodology, query or suggestion
   0 ALERT type 5 Informative message, check
```

Figure 8.13 CheckCIF report from fraudulent synthetic data, created in September 2018.

Since this potential method of fraud was highlighted, a number of additional checks have been added to the *CheckCIF/Platon* software,[238] as visible in an updated CheckCIF report (Figure 8.13) including a test for positive residual density between C – C bonds (PLAT978).[213] It would be expected that a number of the C – C bonds show positive residual electron density after IAM refinement – as was present in the IAM in Chapters 3-6. In the reported

fraudulent structure, because it was created using the calculated structure factors with little-to-no noise included (and no residual density values), the final model would be very similar to the original and therefore not contain any residual electron density.

This work also led to the addition of a check for negative residual electron density at the location of the hydrogen atom (PLAT977). As presented by Dr. Spek, as a consequence of changing the structure and removing part of the five-membered ring, the CH_2 hydrogen atoms have negative density surrounding then, indicating they were in an incorrect position (Figure 8.14).

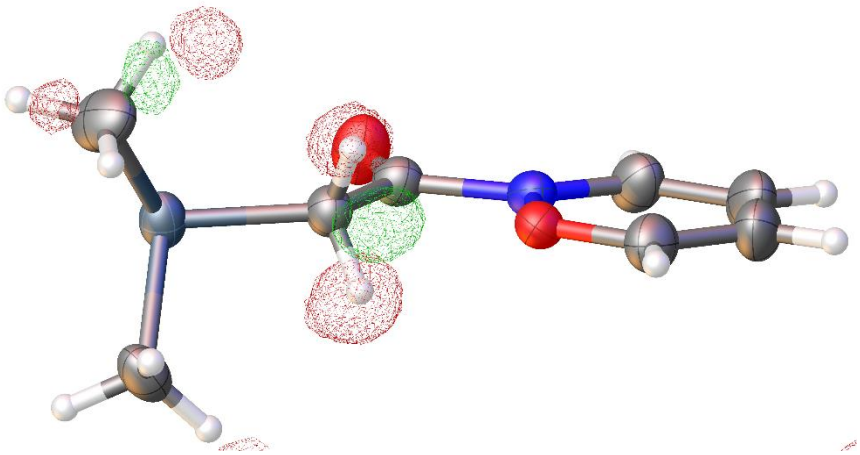


Figure 8.14 Residual electron density plot calculated with Olex2 for fraudulent ylid structure at ± 0.1 eÅ^{-3} level. The negative density around the hydrogen atoms (red), along with the positive density between the atoms (green) indicated the hydrogen atoms have been incorrectly placed.

Dr. Spek states that analysis of the fraudulent structure, even before the additional checks were added, would have raised cause for concern due to the negligible residual densities within the model.[239] A test in *PLATON* already existed to look for the presence of negative residual density (PLAT961), which was added since the initial testing and would also highlight the current lack of residual density in the structure due to it being produced using a spherical atom model. However, if the data were replicated using structure factors from a model containing aspherical densities, this could make identification of fraud more difficult. The additional density in the model from the aspherical densities would result in residual electron density being present between atoms when modelling the fraudulent structure using the spherical atom model. Models which contain information from aspherical densities could easily be created using multipolar densities from one of the many TAAM databanks.

In order to safeguard diffraction data, it will be the role of the instrument manufacturers to include validity checks in their software. Without user demand for this, there may be little reason for the manufacturers to include such security measures in their software. It is

impossible to defend against the most determined fraudster, but with significant hurdles and checks it should be as hard as possible an act to get away with. As a community, crystallographers must be constantly vigilant to potential future avenues of fraud and put in place measures to stop those avenues from being utilised to protect the integrity of the technique.

Chapter 9. Conclusions

Throughout the course of this thesis, a number of optimisation methods for the processing of charge density data have been presented – including a method for the calculation and application of the optimised values for the *a* and *b* parameters of a *SHELXL* weighting scheme. Additionally, an intuitive GUI for the easy operation of the program to calculate these parameters and provide normal probability plots of diffraction data residuals has been developed and released. Its use has been invaluable in the diagnosis of certain issues with high resolution X-ray data.

The effects of several different processing options were investigated to assess the effect on the multipolar model and the outcome of the refinement. In Chapter 3, only chemically constraining atoms produced statistically different multipolar parameters from the reference model. This suggests that the multipolar model derived from X-ray diffraction data is robust and resistant to data processing options, at least in terms of global features. It also suggests that the chemical similarity constraints available in *XD2016* may not always be applicable for the same chemical environments, despite their use to prevent overfitting. The cause of the differences in multipolar parameters for seemingly chemically equivalent atoms within the molecule is currently unknown. All other processing choices investigated showed no statistical difference in multipolar parameters or other properties calculated. However, the order of inclusion of multipolar parameters was found to be critical to allow convergence of multipole parameters and therefore should always be considered carefully.

Additionally, multipolar parameters in chemically equivalent environments within different data studies were all found to be generally transferable when investigated using polymorphs of the same compound (Chapter 4) and with a structure of Z' = 2 (Chapter 5). There was also a general amount of transferability between chemically equivalent atoms in TCNQ and DCMC (Chapter 6). This is very encouraging as this property is key to TAAM, a form of modelling which is becoming more prevalent in chemistry and will allow the expansion of the technique to ever more complex systems.

The analysis of all residuals within the normal probability plot (using *CAPOW*) is key in identifying the presence of systematic errors within diffraction data. The advantages of the analysis of the full range of the residuals in a normal probability plot has been highlighted in the identification of large residual values, which are characteristic of problems within the dataset. The facility to identify the reflections by Miller indices within the normal probability plot created by *CAPOW* allowed investigation into problematic reflections to allow the diagnosis of systematic errors. Classical analysis of the residuals within the range of ± 4 would not have been sufficient to identify errors, such as white radiation contamination.

White radiation contamination was shown to be an issue within the multipolar refinement field when using multilayer optics. This source of systematic error has been identified within a number of datasets by the occurrence of large residual values within normal probability plots, even after the *SHELXL* weighting scheme was applied. Careful analysis is required to assess for the presence of this effect, as for L-ascorbic acid (Chapter 5) the contamination was not immediately obvious in the original calculated precession images; the creation of precession images with short scan lengths included can hide the characteristics of the error. The prevention or correction of this contamination is imperative for the production of reliable results; however, the current corrections of white radiation are not applicable to Ag X-ray sources. The prevalence of white radiation could suggest that the longer exposure times which are utilised in charge density in order to reduce the signal to noise ratio for data can make this effect more prevalent than when collecting data for spherical atom refinement

The underestimation of standard uncertainties within the multipolar refinements is a known issue and is seen in the statistically weighted data throughout this thesis. The application of a *SHELXL* weighting scheme has been shown to produce an approximately normal distribution and reduce the range of residuals to ±4. If this does not occur, systematic errors, such as white radiation contamination, may still be present in the data that require attention before the refinement is considered complete (Chapter 5). The application of a *SHELXL* weighting scheme has been shown to reduce the range of residual electron density and other associated parameters. However, there was no systematic difference between multipolar parameters calculated when using a statistical or a *SHELXL* weighting scheme, with only small systematic differences in the Laplacian values at some BCPs. This may be a result of the low standard uncertainties calculated, as not all sources of error (e.g. errors in

position and U_{ij}) are propagated to the final parameters within *XD* and therefore the calculation of standard uncertainties should be larger. Provided there are no sources of systematic error within the model, it is deemed appropriate to apply the *SHELXL* weighting scheme in these datasets – although assessment of both the statistical and *SHELXL* weighted model is needed to ensure no spurious effects are found due to the weighting.

The experimental aim is always to produce the best model fit to the observed data, however, there are still meaningful results that can be gained from models with higher residual electron densities – especially when larger atoms are present, where higher electron densities can occur due to the accuracy of the atomic structure factors.

Overall, the results shown in this thesis highlight the promising transferability of multipolar parameters. The test case indicates differences in processing do not produce systematic differences in the multipole parameters. However, there are more sensitive parameters which were not compared, e.g. dipole moments. It is to be expected that these more sensitive derived parameters would show a larger disparity with different data collection or processing method. Analysis of the transferability of properties calculated from the multipole model was limited to properties at BCPs and net atomic charges as *XD2016* only provides standard uncertainty values for certain calculated parameters. This may not be the case for more complex datasets or data which require specific corrections that are critical to achieving a normal distribution of residuals.

Chapter 10. Appendices

10.1 Additional information for Chapter 3

10.1.1 Data frame showing ice contamination

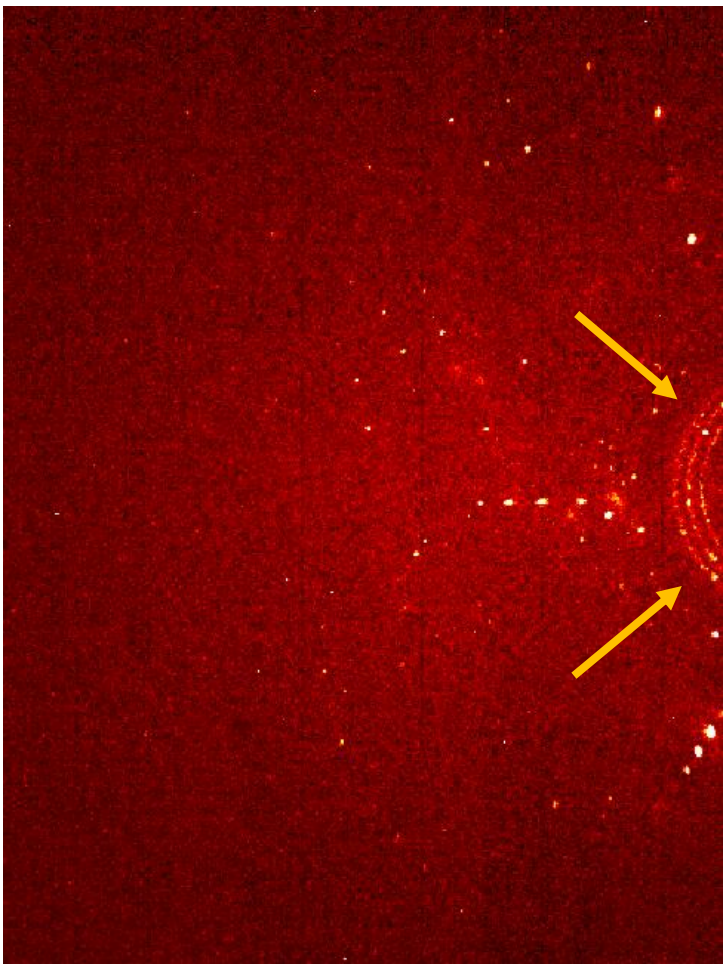


Figure 10.1 Frame 270 Run 22. Ice rings are indicated with arrows.

The rings present in Figure 10.1 are characteristic of ice being present on crystal. Ice grows in many single crystals, therefore creates powder-diffraction-esque diffraction pattern. Ice rings will affect any reflection which falls within the rings (as additional diffraction can make reflections appear more intense) therefore causing higher intensities to be measured. The extent of these rings is 1.99 Å.

10.1.2 Data collection strategy for reference data collection of F_4TCNQ

Run	Scan	2theta(°)	Omega(°)	Phi(°)	Scan length (s)	Scan width	# of frames
1	Phi	0.0	0.0	0.0	1	1	180
2	Omega	53.78	225.34	48.00	39	0.5	268
3	Omega	53.78	286.00	344.67	39	0.5	146
4	Omega	53.78	225.34	69.85	39	0.5	268
5	Phi	35.37	43.80	71.38	20	0.5	338
6	Omega	53.78	225.34	247.81	39	0.5	268
7	Omega	53.78	256.00	289.67	39	0.5	186
8	Omega	53.78	225.34	41.39	39	0.5	268
9	Omega	53.78	266.00	129.35	39	0.5	186
10	Phi	35.37	206.93	215.38	20	0.5	290
11	Omega	53.78	225.34	30.45	39	0.5	268
12	Omega	53.78	266.00	307.34	39	0.5	186
13	Omega	53.78	225.34	223.07	39	0.5	268
14	Omega	14.80	186.36	69.98	11	0.5	346
15	Omega	53.78	276.00	322.11	39	0.5	166
16	Omega	53.78	225.34	34.63	39	0.5	268
17	Omega	8.06	179.63	250.32	10	0.5	359
18	Omega	53.78	225.34	207.48	39	0.5	268
19	Omega	54.78	306.00	187.73	39	0.5	106
20	Omega	6.62	226.00	322.11	10	0.5	196
21	Omega	53.78	225.34	265.14	39	0.5	228
22*	Phi	54.68	63.11	279.49	40	0.5	255 (466)
23*	Omega	53.78	225.34	87.52	39	0.5	227

*.Table 10.1 Data collection strategy for F_4TCNQ *data from run 22, frame 255 were removed from final processing due to presence of ice on crystal.*

10.1.3 Multipolar refinement strategies

10.1.3.1 Reference multipolar refinement strategy

Key: Monopoles, dipoles, quadrupoles, octupoles and hexadecapoles denoted by the initial (M, D, Q, O, H). κ and κ' are the kappa parameters.

Setup refinement - as *XDINI* does not carry over all Uijs from *SHELX* ins file, therefore these values must be refined in XD.

1. Scale only, statistical weights
2. xyz
3. xyz and Uij
4. xyz (sinθ/λ 0.8 to -1.2)
5. xyz, Uij (sinθ/λ 0.8 to -1.2)
6. scale only (sinθ/λ 0.8 to -1.2)
7. xyz, Uij (sinθ/λ 0.8 to -1.2)

Multipolar refinement – no sinθ cut-off, I/σ < 3 removed, statistical weights.

1. M
2. D, Q, O, H
3. M, D, Q, O, H
4. Uij
5. M, D, Q, O, H
6. Uij, M, D, Q, O, H
7. xyz
8. xyz, M, D, Q, O, H
9. xyz, Uij, M, D, Q, O, H
10. κ, xyz, Uij, M, D, Q, O, H
11. M
12. κ, M
13. xyz, Uij, M, D, Q, O, H
14. κ
15. xyz, Uij, M, D, Q, O, H
16. κ, xyz, Uij, M, D, Q, O, H
17. xyz, Uij, M, D, Q, O, H
18. κ '
19. D, Q, O, H
20. κ, xyz, Uij, M, D, Q, O, H
21. D, Q, O, H
22. xyz, Uij, M, D, Q, O, H
23. κ, xyz, Uij, M, D, Q, O, H

10.1.3.2 Refinement strategy: κ' refined twice

Multipolar refinement – no sinθ/λ cut-off, I/σ < 3 removed, statistical weights.

1. M
2. D, Q, O, H
3. M, D, Q, O, H
4. Uij
5. M, D, Q, O, H
6. Uij, M, D, Q, O, H
7. xyz
8. xyz, M, D, Q, O, H
9. xyz, Uij, M, D, Q, O, H
10. xyz, Uij, M, D, Q, O, H, κ

11. M
12. κ M
13. xyz, Uij, M, D, Q, O, H
14. κ
15. xyz, Uij, M, D, Q, O, H
16. K, xyz, Uij, M, D, Q, O, H
17. xyz, Uij, M, D, Q, O, H
18. κ '
19. D, Q, O, H
20. κ, xyz, Uij, M, D, Q, O, H
21. κ '
22. D, Q, O, H
23. xyz, Uij, M, D, Q, O, H
24. κ, xyz, Uij, M, D, Q, O, H

If cut-offs are removed

25. xyz, Uij, M, D, Q, O, H, I/σ_I <2
26. xyz, Uij, M, D, Q, O, H, I/σ_I <1
27. xyz, Uij, M, D, Q, O, H, no I/σ_I cut-off
28. κ, xyz, Uij, M, D, Q, O, H

10.1.3.3 Refinement strategy: constrained/unconstrained refinement

Multipolar refinement – no sinθ/λ cut-off. I/σ < 2 removed, statistical weights.

1. M
2. D, Q, O, H
3. M, D, Q, O, H
4. Uij
5. M, D, Q, O, H
6. Uij, M, D, Q, O, H
7. xyz
8. xyz, M, D, Q, O, H
9. xyz, Uij, M, D, Q, O, H
10. κ
11. M
12. κ, M
13. xyz, Uij, M, D, Q, O, H
14. κ, xyz, Uij, D, Q, O, H
15. M
16. xyz, Uij, M, D, Q, O, H
17. κ
18. xyz, Uij, M, D, Q, O, H
19. κ, xyz, Uij, M, D, Q, O, H
20. xyz, Uij, D, Q, O, H
21. κ '
22. D, Q, O, H
23. κ, xyz, Uij, M, D, Q, O, H
24. κ '
25. D, Q, O, H
26. xyz, Uij, M, D, Q, O, H
27. κ, xyz, Uij, M, D, Q, O, H

10.1.4 Additional plots for Section 3.4.1

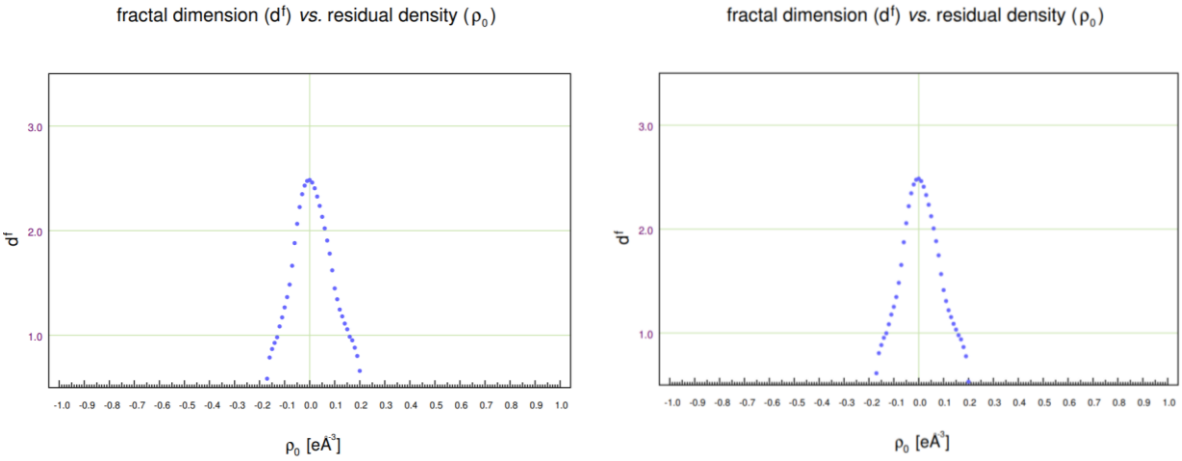


Figure 10.2 Residual density analysis for no κ' refined (left) and κ' refined twice (right).

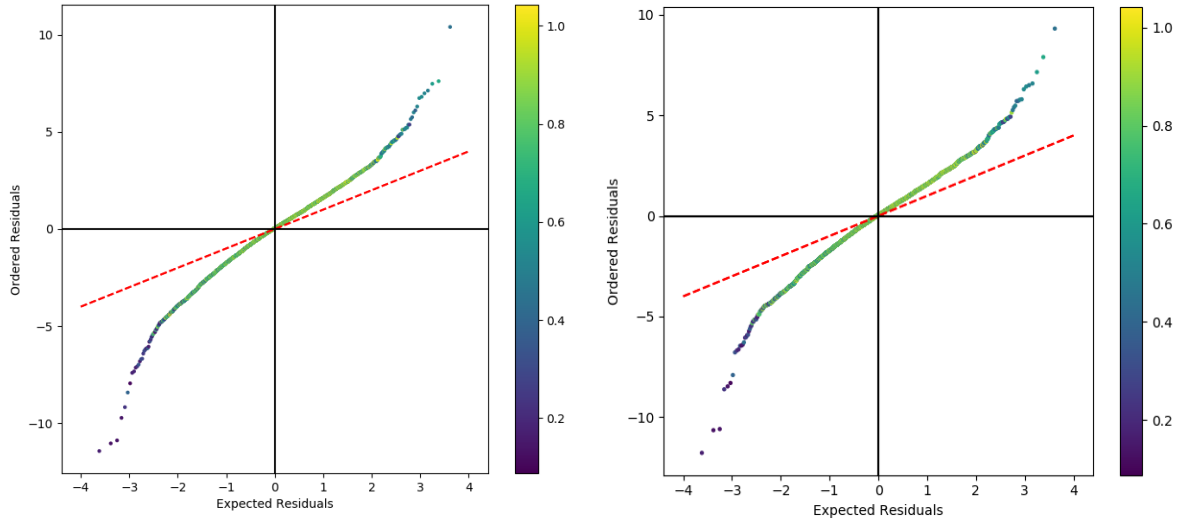


Figure 10.3 Normal probability plot for no κ' refined (left) and κ' refined twice (right). Colourbar shows the resolution of the reflection in Å^{-1}.

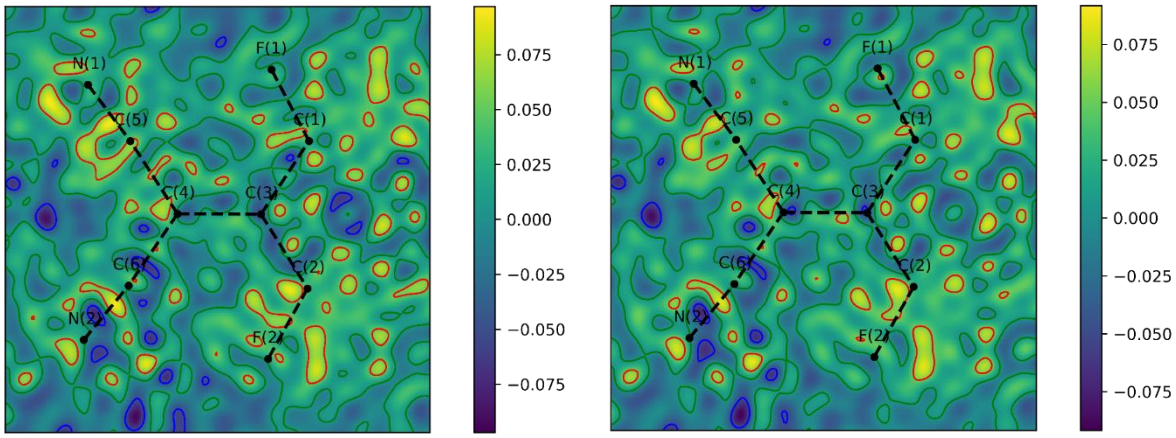


Figure 10.4 Residual electron density plot in the plane of the molecule for no κ' refined (left) and κ' refined twice (right). Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

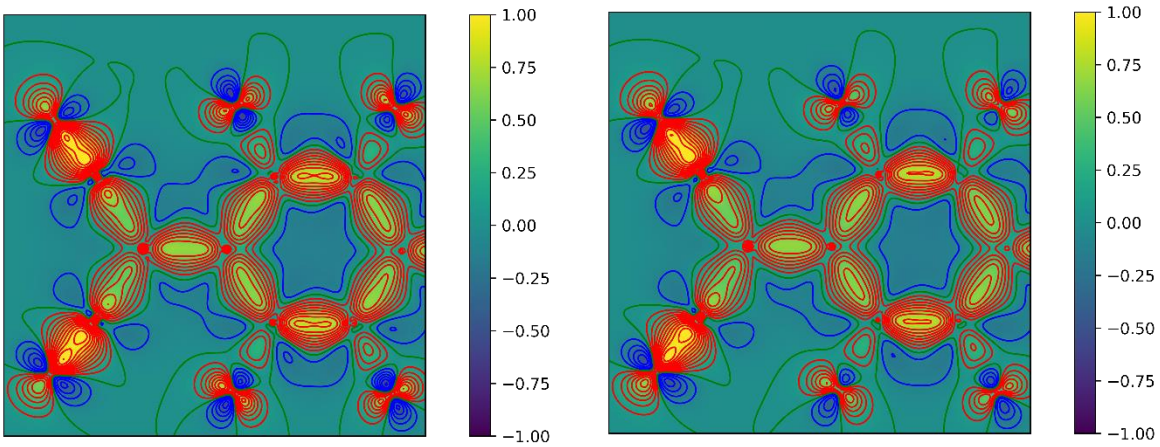


Figure 10.5 Static deformation electron density for no κ' refined (left) and κ' refined twice (right). Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

10.1.5 Additional plots for Section 3.4.2

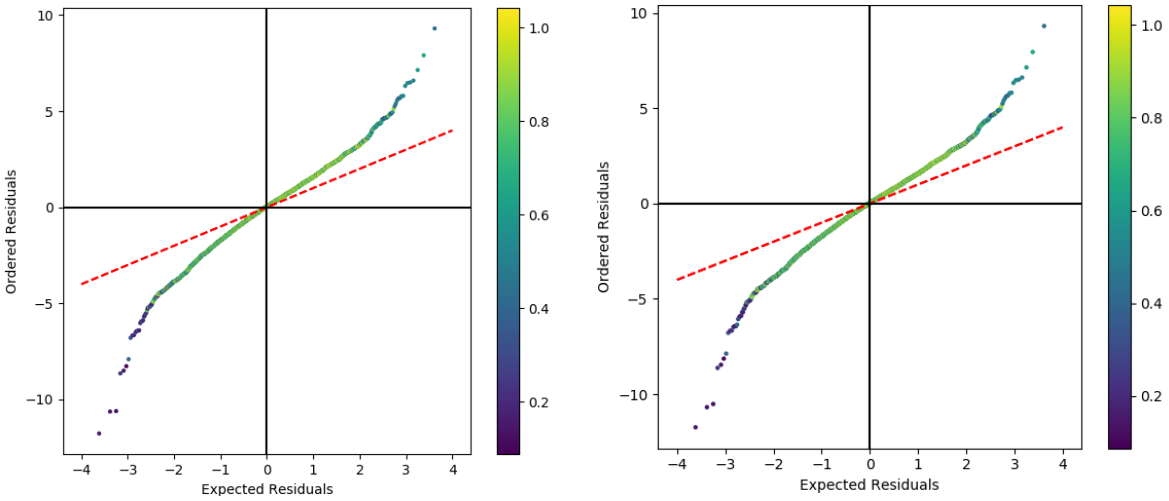


Figure 10.6 Normal probability plot for no $I/\sigma_I < 2$ cut-off (left) and $I/\sigma_I < 0$ cut-off (right). The colourbar shows the resolution of the reflection in $Å^{-1}$.

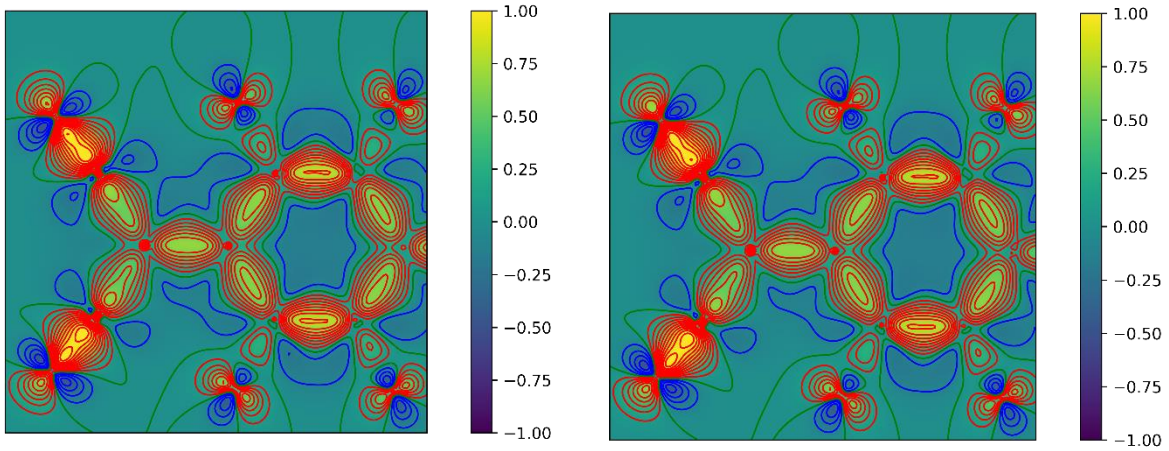


Figure 10.7 Static deformation electron density for no $I/\sigma_I < 2$ cut-off (left) and $I/\sigma_I < 0$ cut-off (right). Contour values are drawn at levels of ± 0.1 $eÅ^{-3}$, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 $eÅ^{-3}$ in order to better view variation of values across the plot.

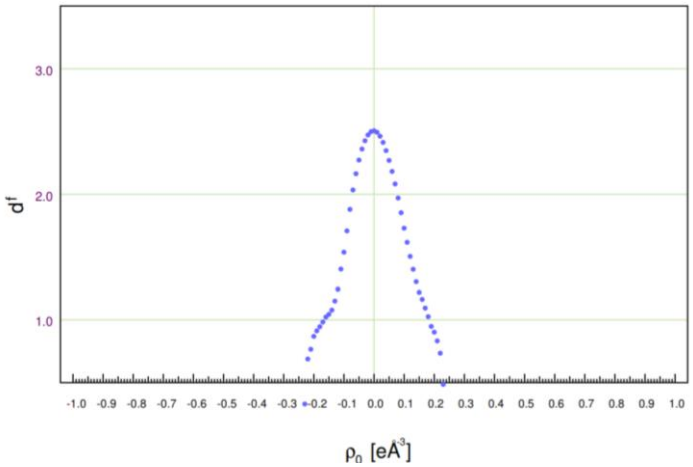


Figure 10.8 Fractal dimensionality plot for multipole refinement with $I/\sigma_I < 0$ cut-off.

10.1.6 Additional plots for Section 3.4.3

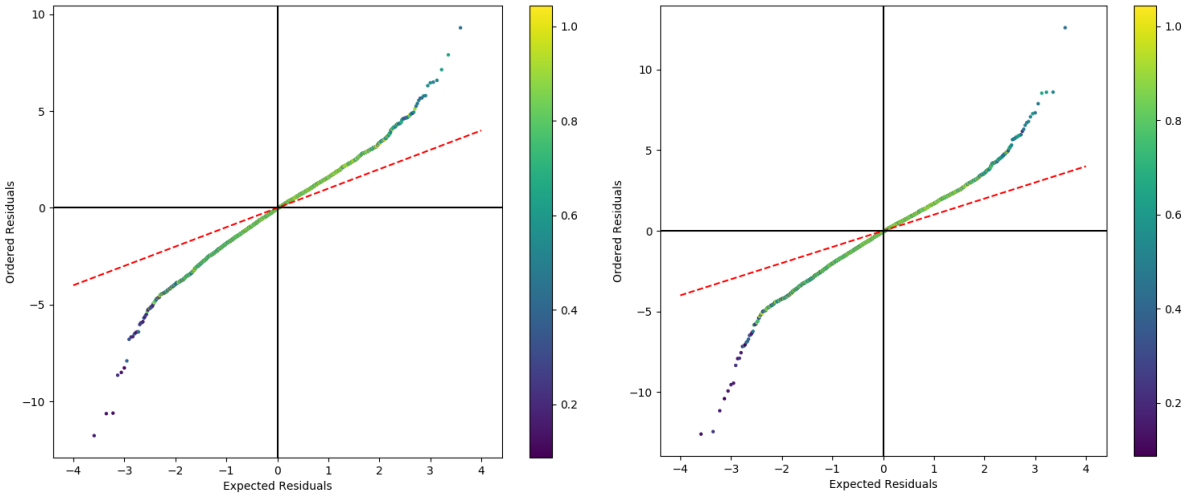


Figure 10.9 Normal probability plot for unconstrained (left) and chemically constrained (right) multipole refinement. Colourbar shows the resolution of the reflection in Å^{-1}.

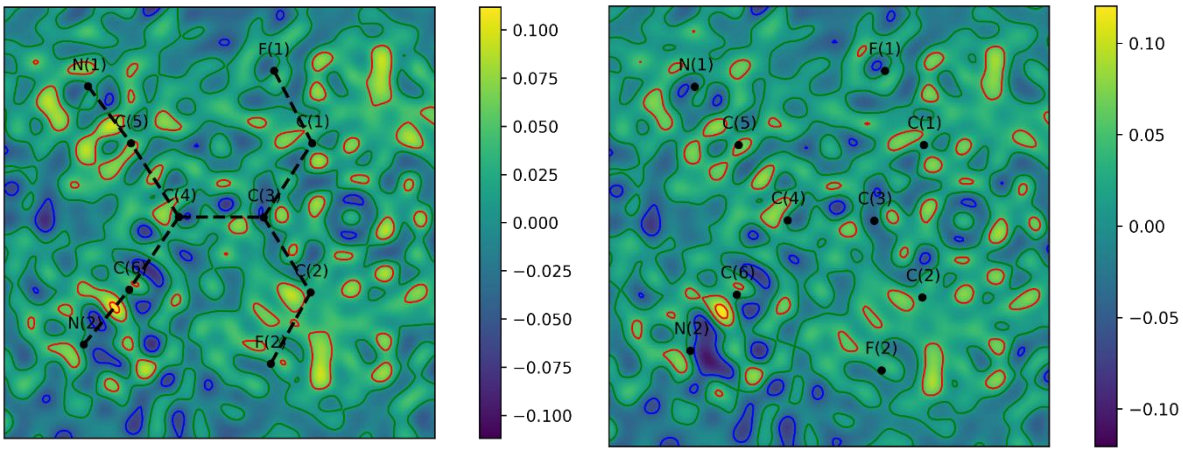


Figure 10.10 Residual electron density plots in the plane of the molecule for unconstrained (left) and chemically constrained (right) multipole refinement. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

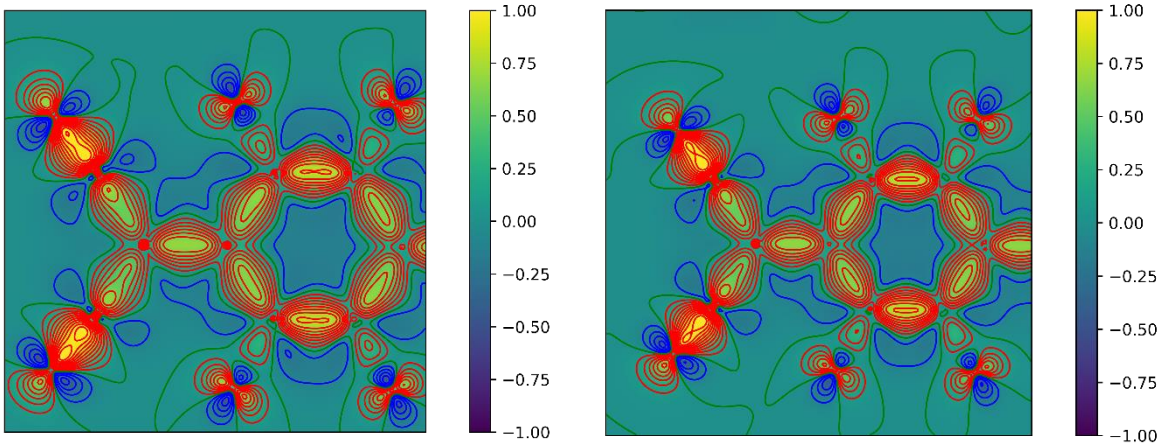


Figure 10.11 Static deformation electron density plots for unconstrained (left) and chemically constrained (right) multipole refinement. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

10.1.7 Additional plots for Section 3.4.4

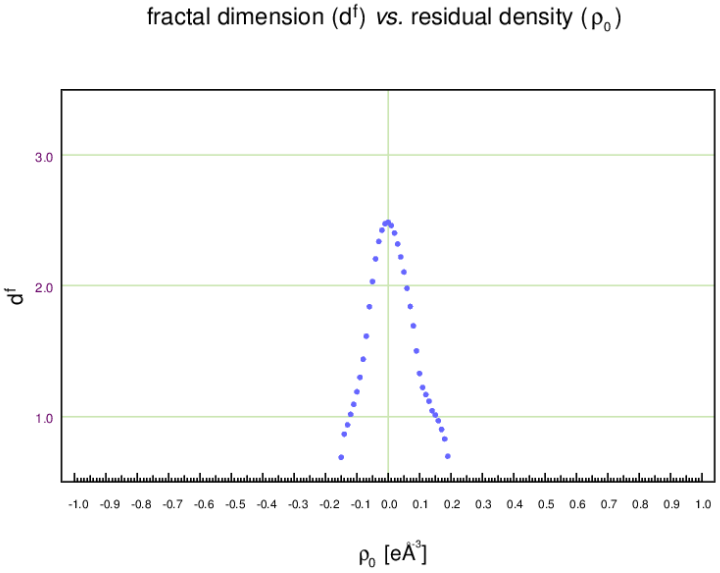


Figure 10.12 Fractal dimensionality plot for SHELXL weighted refinement.

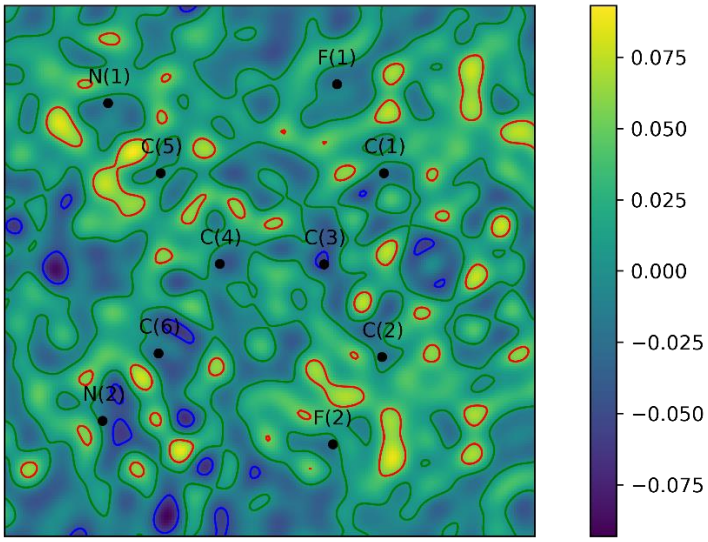


Figure 10.13 Residual electron density plot in the plane of the molecule for the SHELXL weighted multipole refinement. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

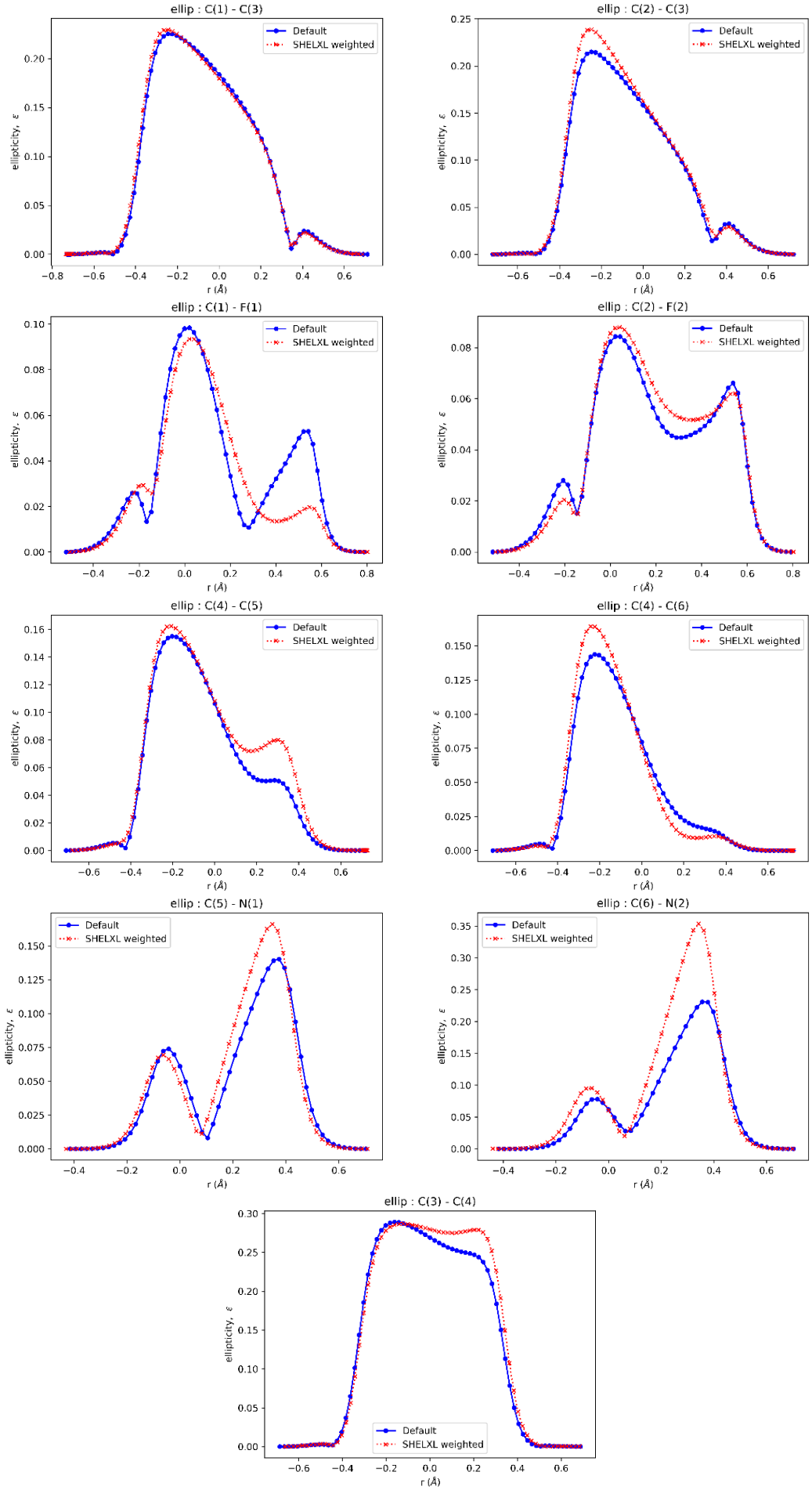


Figure 10.14 Ellipticity along the bond path for bonds within F_4TCNQ of the SHELXL weighted refinement.

10.1.8 Additional plots and tables for Section 3.5.1

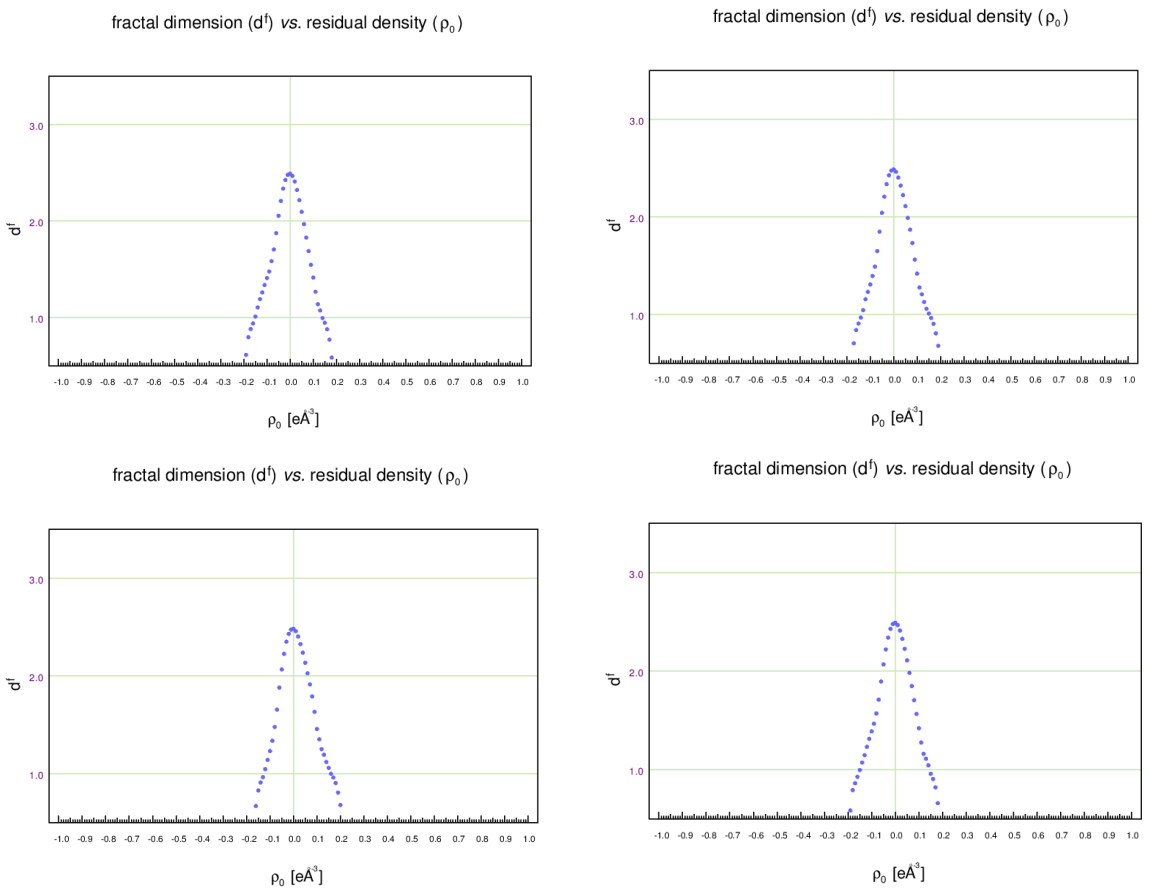


Figure 10.15 Fractal dimensionality plots for error models 0 (top left), 1 (top right), 6 (bottom left) and 8 (bottom right).

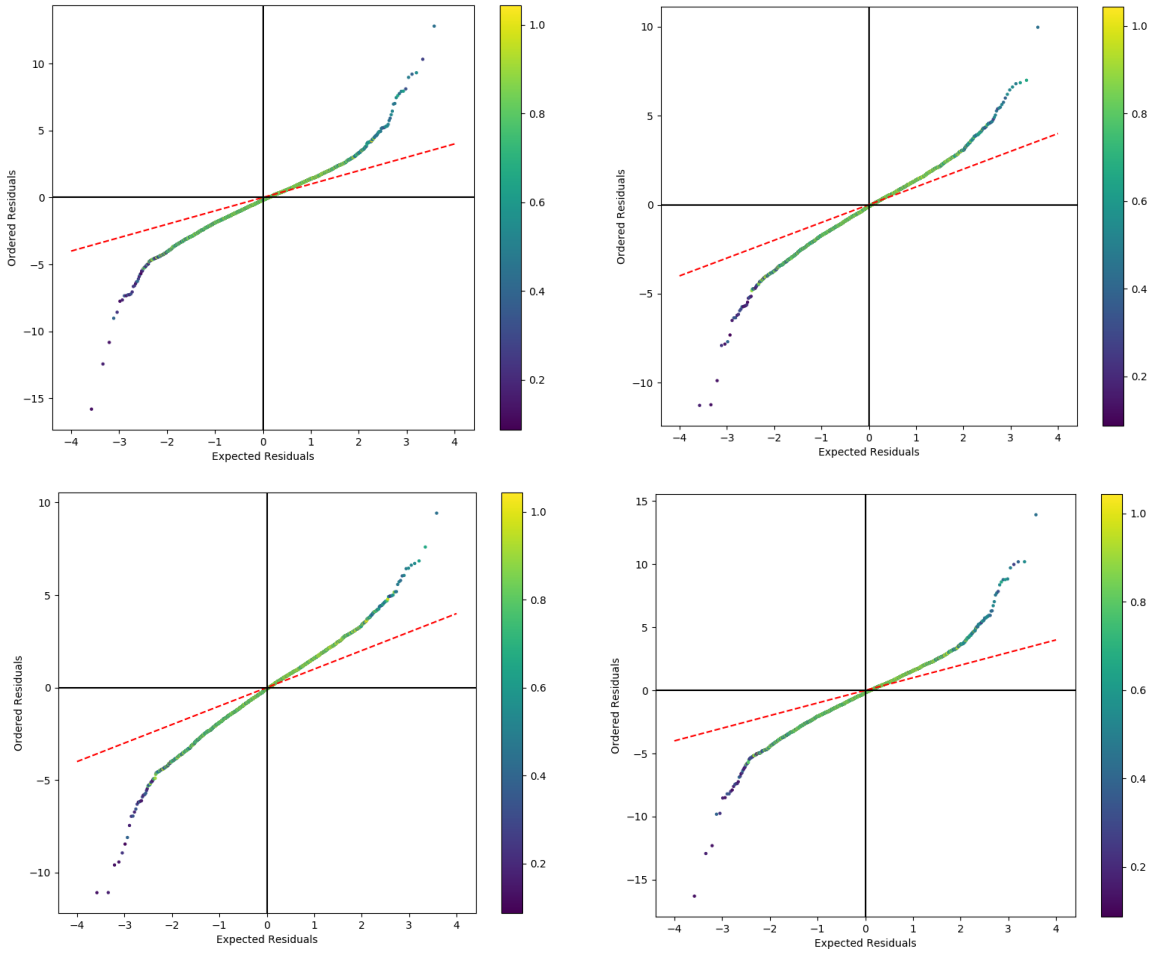


Figure 10.16 Normal probability plots for error models 0 (top left), 1 (top right), 6 (bottom left) and 8 (bottom right). Colourbar shows the resolution of the reflection in Å^{-1}.

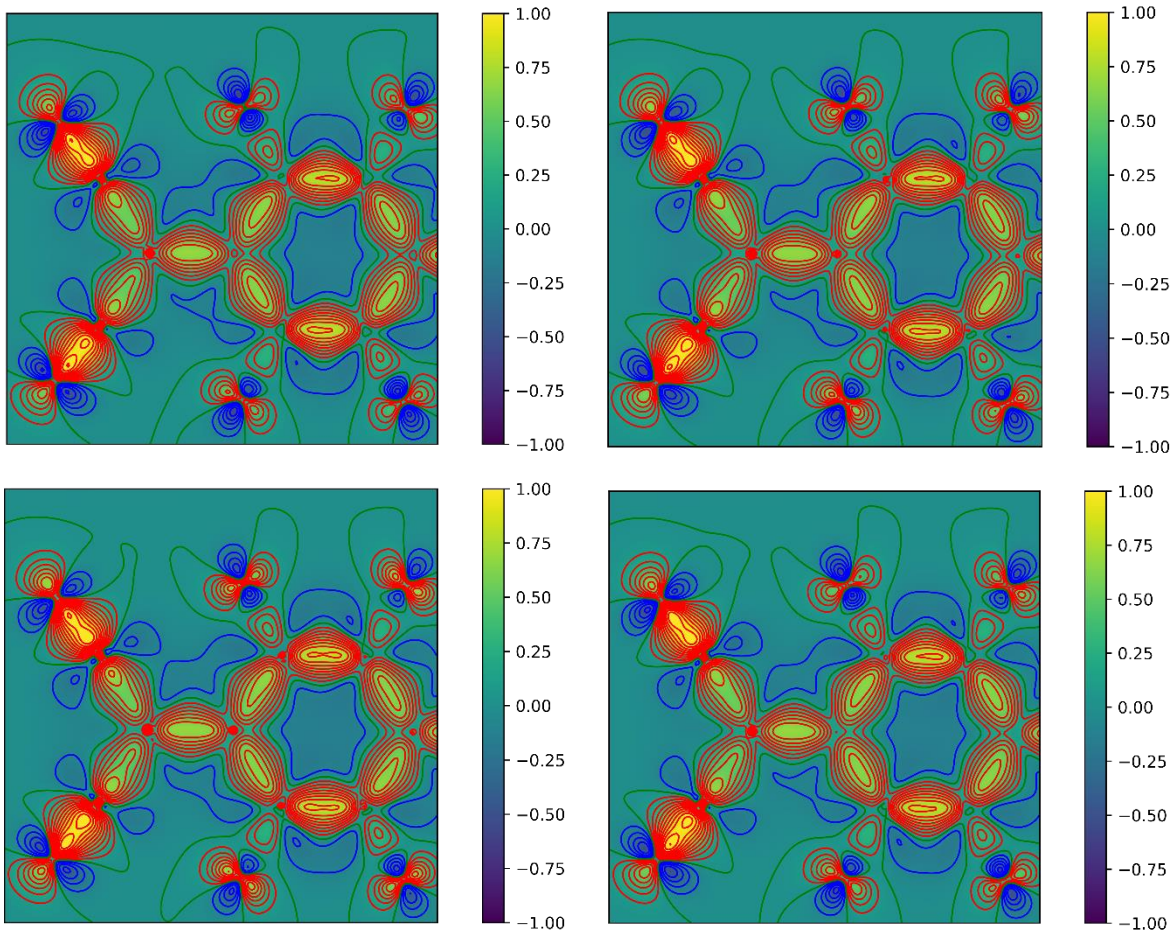


Figure 10.17 Static deformation electron density plots for error models 0 (top left), 1 (top right), 6 (bottom left) and 8 (bottom right). Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

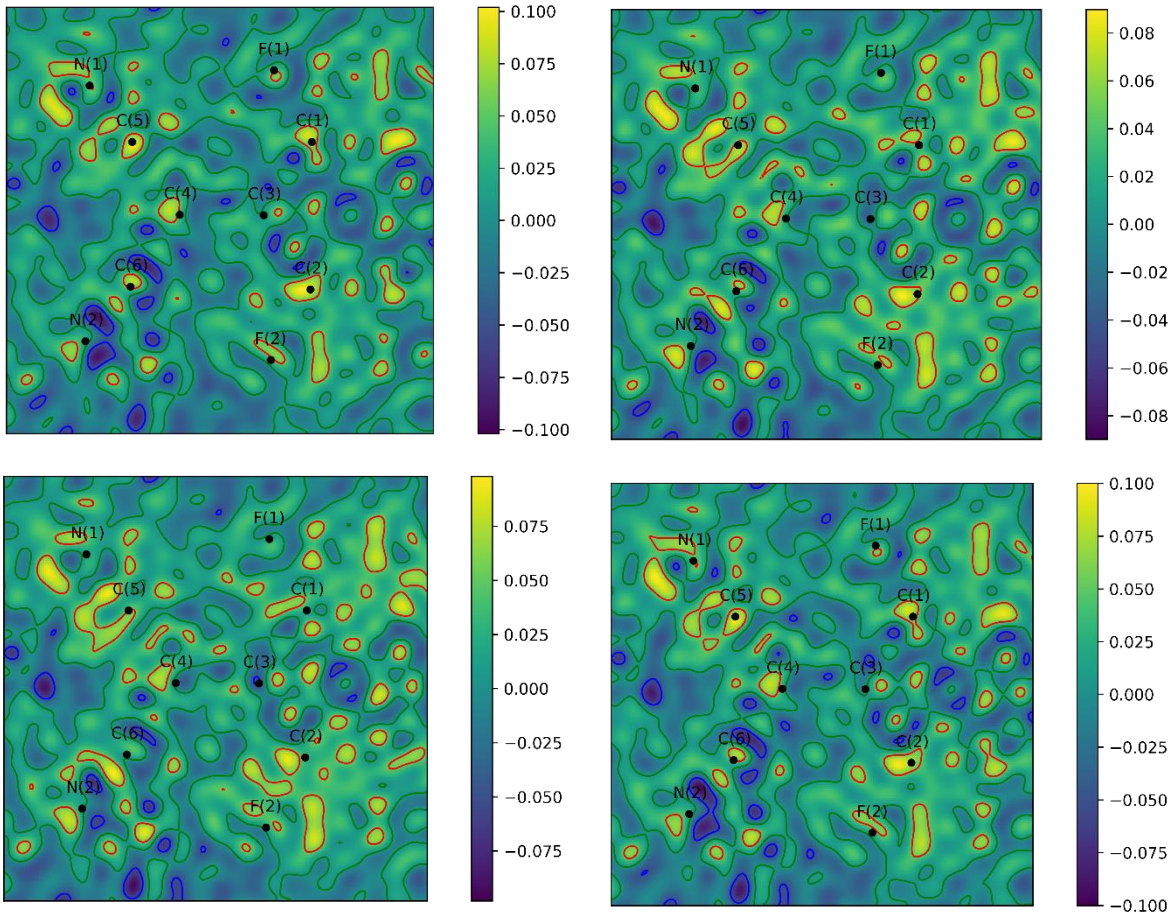


Figure 10.18 Residual electron density plots in the plane of the molecule for error models 0 (top left), 1 (top right), 6 (bottom left) and 8 (bottom right). Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

	Error 0				Error 1			
Atom	P_{val}	κ	κ′	Net atomic charge	P_{val}	κ	κ′	Net atomic charge
F(1)	7.067(10)	0.995	0.937	-0.066(10)	7.091(11)	0.998	0.947	-0.090(11)
F(2)	7.064(10)	0.995	0.937	-0.063(10)	7.087(10)	0.998	0.947	-0.087(10)
N(1)	5.104(24)	0.991(1)	1.006	-0.104(24)	5.124(25)	0.994(2)	1.008	-0.124(25)
N(2)	5.068(25)	0.991(1)	1.006	-0.067(25)	5.074(25)	0.994(2)	1.008	-0.073(25)
C(1)	3.918(19)	1.008(2)	0.98	+0.081(19)	3.909(19)	1.014(2)	0.984	+0.091(19)
C(2)	3.902(19)	1.008(2)	0.98	+0.098(19)	3.888(19)	1.014(2)	0.984	+0.111(19)
C(3)	3.988(21)	1.000(2)	0.974	+0.012(21)	3.986(21)	1.007(2)	0.978	+0.014(21)
C(4)	4.123(22)	1.000(2)	0.974	-0.122(22)	4.095(22)	1.007(2)	0.978	-0.095(22)
C(5)	3.849(28)	1.032(2)	0.989	+0.150(28)	3.839(28)	1.038(2)	0.989	+0.160(28)
C(6)	3.918(28)	1.032(2)	0.989	+0.082(28)	3.908(29)	1.038(2)	0.989	+0.092(29)
	Error 4				Error 5			
Atom	P_{val}	κ	κ′	Net atomic charge	P_{val}	κ	κ′	Net atomic charge
F(1)	7.110(11)	0.999	0.954	-0.110(11)	7.075(10)	0.996	0.941	-0.075(10)
F(2)	7.104(11)	0.999	0.954	-0.104(11)	7.071(10)	0.996	0.941	-0.071(10)
N(1)	5.107(25)	0.997(2)	1.01	-0.106(25)	5.092(24)	0.993(1)	1.006	-0.091(24)
N(2)	5.057(26)	0.997(2)	1.01	-0.056(26)	5.057(25)	0.993(1)	1.006	-0.057(25)
C(1)	3.907(19)	1.017(2)	0.985	+0.092(19)	3.914(19)	1.010(2)	0.981	+0.086(19)
C(2)	3.880(19)	1.017(2)	0.985	+0.119(19)	3.897(19)	1.010(2)	0.981	+0.103(19)
C(3)	3.974(22)	1.012(2)	0.98	+0.026(22)	3.993(21)	1.002(2)	0.975	+0.007(21)
C(4)	4.074(22)	1.012(2)	0.98	-0.074(22)	4.124(22)	1.002(2)	0.975	-0.124(22)
C(5)	3.862(29)	1.041(3)	0.99	+0.138(29)	3.854(28)	1.034(2)	0.989	+0.146(28)
C(6)	3.925(29)	1.041(3)	0.99	+0.075(29)	3.923(29)	1.034(2)	0.989	+0.076(29)

Table 10.2 Net atomic charges for multipole refinements from data refined with different error models.

		Error model 0			Error model 1		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.134(16)	-24.501(89)	0.11	2.133(15)	-23.757(86)	0.1
F(2)	C(2)	2.102(16)	-23.158(90)	0.08	2.096(15)	-22.100(87)	0.08
N(1)	C(5)	3.368(26)	-26.048(209)	0.07	3.380(25)	-26.707(204)	0.07
N(2)	C(6)	3.339(26)	-25.996(214)	0.06	3.355(25)	-27.408(208)	0.06
C(1)	C(3)	2.019(8)	-17.244(28)	0.18	2.012(8)	-16.808(29)	0.18
C(2)	C(3)	2.019(8)	-17.765(28)	0.16	2.011(8)	-17.257(29)	0.16
C(1)	C(2*)	2.395(27)	-24.148(67)	0.38	2.392(27)	-23.766(67)	0.39
C(3)	C(4)	2.395(27)	-24.148(67)	0.27	2.212(9)	-20.283(35)	0.27
C(4)	C(5)	1.958(9)	-14.993(30)	0.1	1.956(9)	-14.898(31)	0.1
C(4)	C(6)	1.923(9)	-13.646(30)	0.08	1.915(9)	-13.154(31)	0.08
		Error model 6			Error model 8		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.137(15)	-23.063(84)	0.09	2.141(15)	-24.285(87)	0.1
F(2)	C(2)	2.098(15)	-21.627(86)	0.08	2.107(15)	-23.028(88)	0.08
N(1)	C(5)	3.402(24)	-29.457(197)	0.06	3.381(25)	-27.728(204)	0.07
N(2)	C(6)	3.379(24)	-30.230(199)	0.06	3.351(25)	-27.329(210)	0.06
C(1)	C(3)	2.009(8)	-16.548(30)	0.18	2.018(8)	-17.124(28)	0.18
C(2)	C(3)	2.008(9)	-17.002(30)	0.16	2.019(8)	-17.626(28)	0.16
C(1)	C(2*)	2.392(26)	-23.548(66)	0.39	2.396(27)	-24.074(68)	0.38
C(3)	C(4)	2.208(10)	-19.982(35)	0.27	2.218(9)	-20.751(34)	0.27
C(4)	C(5)	1.955(9)	-14.734(31)	0.11	1.958(9)	-14.897(30)	0.1
C(4)	C(6)	1.911(9)	-12.920(32)	0.08	1.923(9)	-13.537(30)	0.08

Table 10.3 Properties at BCPs for multipole refinements from data refined with different error models.

10.1.9 Additional plots and tables for Section 3.5.2

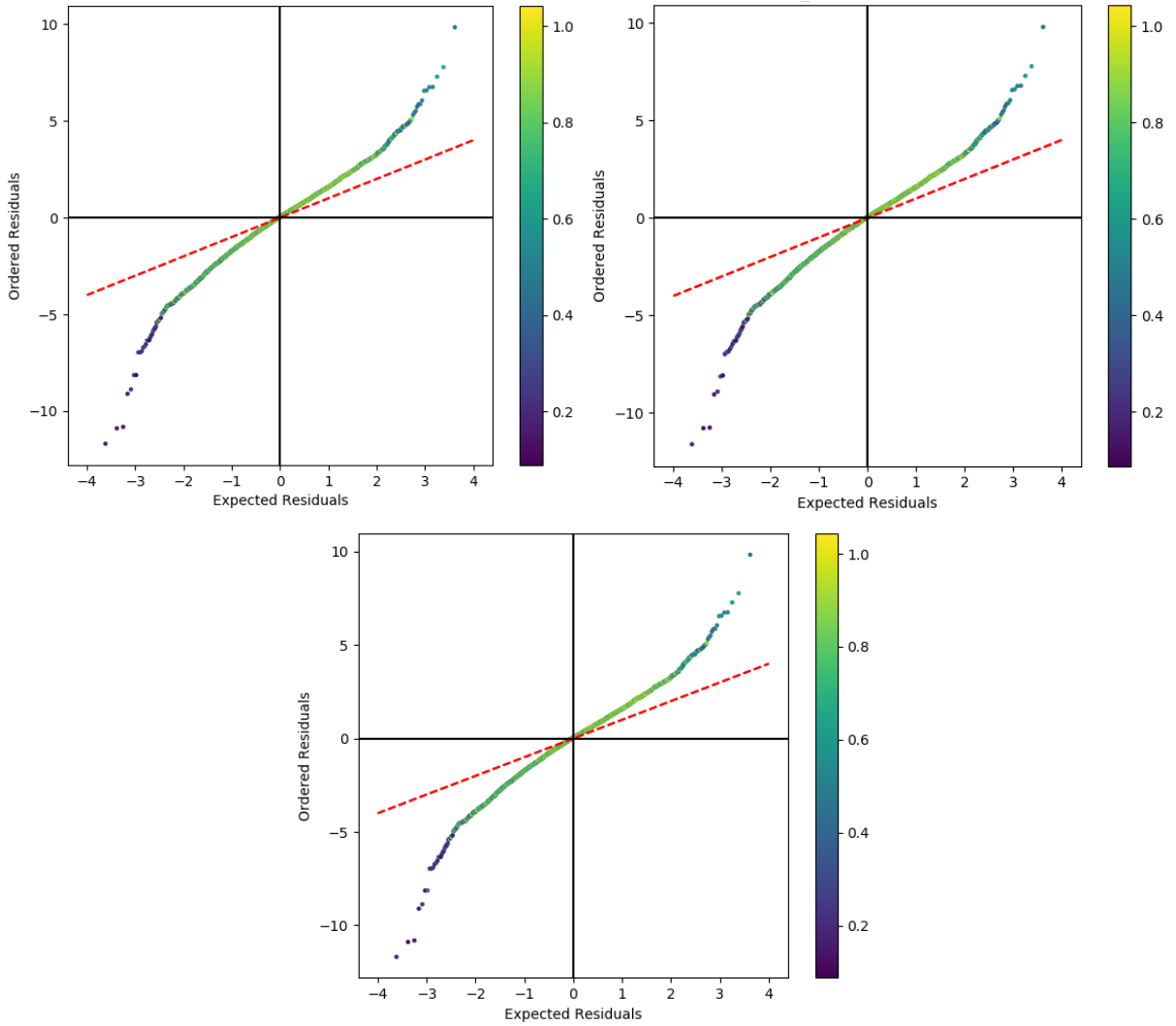


Figure 10.19 Normal probability plots for multipole refinements where no overload correction (top left), no numerical absorption correction (top right), and no overload or numerical absorption correction (bottom) are applied. The colourbar shows the resolution of the reflection in Å^{-1}.

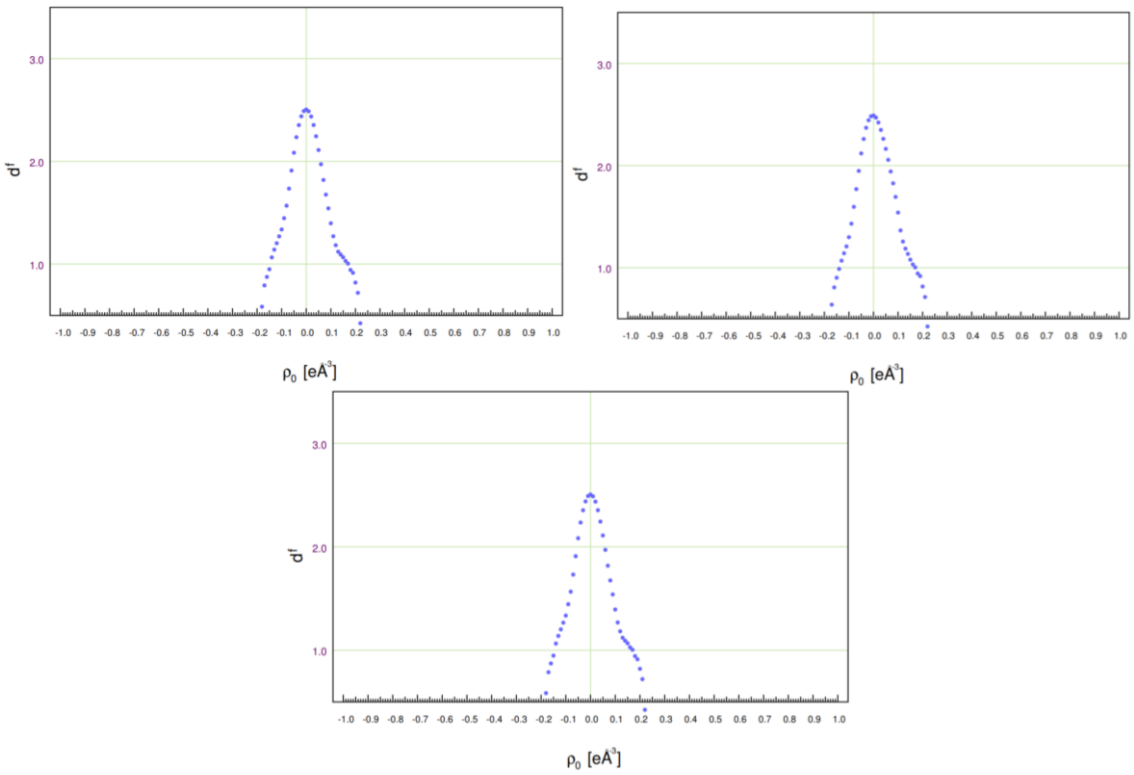


Figure 10.20 Fractal dimensionality plots for multipole refinements where no overload correction (top left), no numerical absorption correction (top right), and no overload or numerical absorption correction (bottom) are applied.

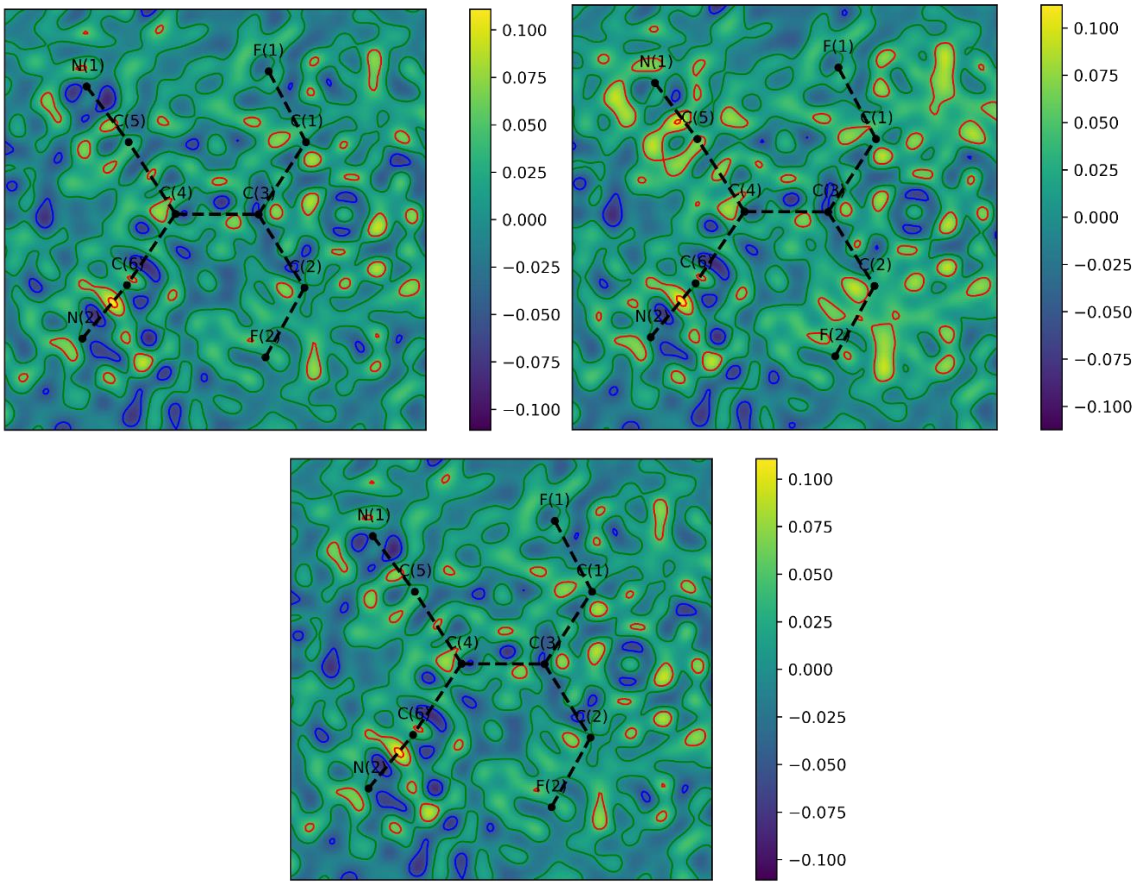


Figure 10.21 Residual electron density plots in the plane of the molecule for multipole refinements where no overload correction (top left), no numerical absorption correction (top right), and no overload or numerical absorption correction (bottom) are applied. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

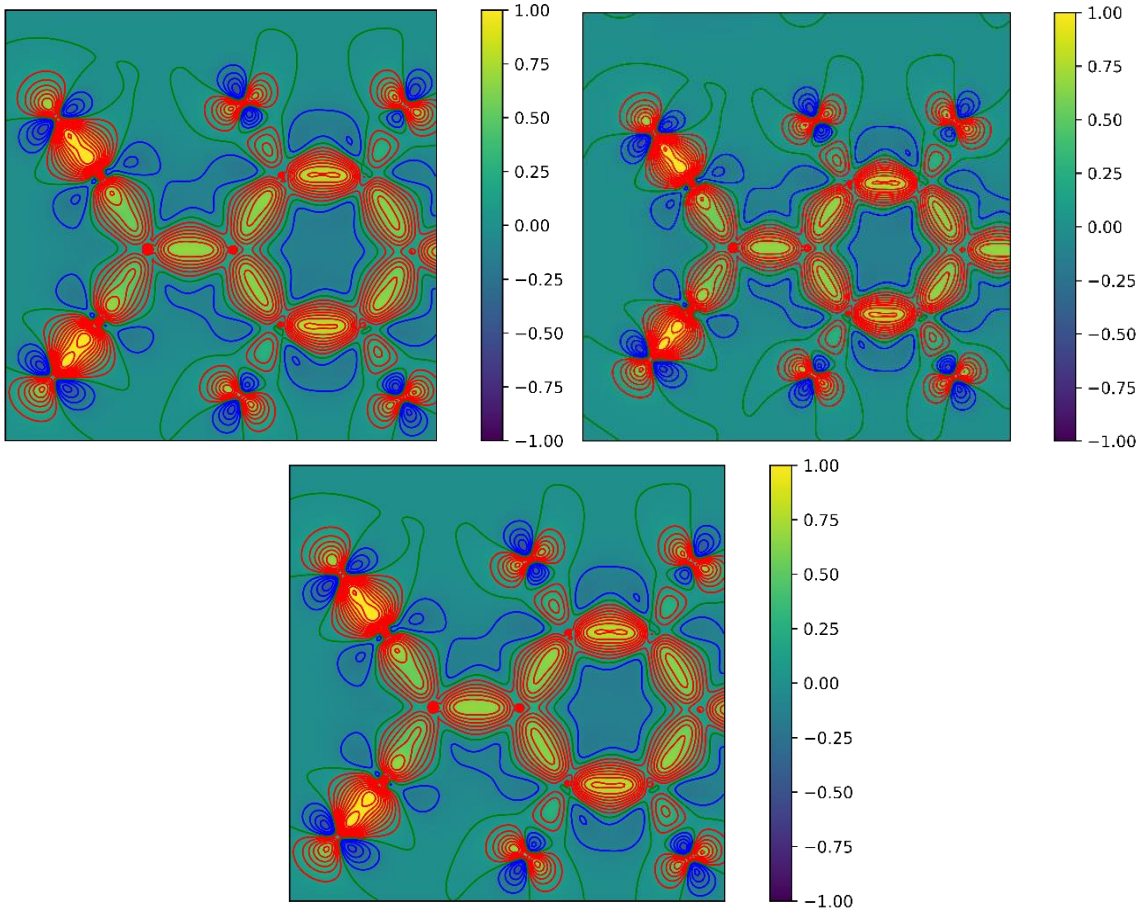


Figure 10.22 Static deformation electron density plots in the plane of the molecule for multipole refinements where no overload correction (top left), no numerical absorption correction (top right), and no overload or numerical absorption correction (bottom) are applied. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

		No overload correction			No numerical correction			No overload or numerical correction		
Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.137(15)	-23.344(83)	0.1	2.136(15)	-23.375(84)	0.1	2.137(15)	-23.346(83)	0.1
F(2)	C(2)	2.099(15)	-21.721(85)	0.08	2.097(15)	-21.725(85)	0.08	2.099(15)	-21.725(85)	0.08
N(1)	C(5)	3.400(24)	-29.257(195)	0.06	3.398(24)	-29.075(195)	0.06	3.399(24)	-29.239(195)	0.06
N(2)	C(6)	3.372(24)	-29.597(198)	0.06	3.372(24)	-29.695(197)	0.06	3.372(24)	-29.602(198)	0.06
C(1)	C(3)	2.010(8)	-16.627(29)	0.18	2.011(8)	-16.659(29)	0.18	2.010(8)	-16.626(29)	0.18
C(2)	C(3)	2.009(8)	-17.108(30)	0.16	2.010(8)	-17.142(30)	0.16	2.009(8)	-17.110(30)	0.16
C(1)	C(2*)	2.391(26)	-23.630(66)	0.39	2.391(26)	-23.632(66)	0.39	2.391(26)	-23.628(66)	0.39
C(3)	C(4)	2.209(9)	-20.055(35)	0.27	2.209(9)	-20.077(35)	0.27	2.209(9)	-20.055(35)	0.27
C(4)	C(5)	1.955(9)	-14.794(31)	0.11	1.954(9)	-14.809(31)	0.11	1.912(9)	-12.963(31)	0.08
C(4)	C(6)	1.912(9)	-12.967(31)	0.08	1.912(9)	-13.005(31)	0.08	1.954(9)	-14.791(31)	0.11

Table 10.4 Properties at BCPs for multipole refinements from data refined for multipole refinements where no overload correction (top left), no numerical absorption correction (top right), and no overload or numerical absorption correction (bottom) are applied.

	No overload correction				No numerical correction				No overload or numerical correction			
Atom	P_{val}	κ	κ′	Net atomic charge	P_{val}	κ	κ′	Net atomic charge	P_{val}	κ	κ′	Net atomic charge
F(1)	7.100(11)	0.999	0.954	-0.100(11)	7.100(11)	0.999	0.953	-0.099(11)	7.100(11)	0.999	0.954	-0.100(11)
F(2)	7.096(11)	0.999	0.954	-0.095(11)	7.095(11)	0.999	0.953	-0.095(11)	7.096(11)	0.999	0.954	-0.095(11)
N(1)	5.105(24)	0.997(2)	1.009	-0.105(24)	5.106(24)	0.996(2)	1.009	-0.105(24)	5.105(24)	0.997(2)	1.009	-0.105(24)
N(2)	5.059(25)	0.997(2)	1.009	-0.058(25)	5.058(25)	0.996(2)	1.009	-0.057(25)	5.059(25)	0.997(2)	1.009	-0.059(25)
C(1)	3.908(19)	1.016(2)	0.984	+0.092(19)	3.908(19)	1.016(2)	0.985	+0.092(19)	3.908(19)	1.016(2)	0.984	+0.092(19)
C(2)	3.882(19)	1.016(2)	0.984	+0.117(19)	3.884(19)	1.016(2)	0.985	+0.116(19)	3.882(19)	1.016(2)	0.984	+0.117(19)
C(3)	3.984(21)	1.010(2)	0.979	+0.016(21)	3.983(21)	1.010(2)	0.979	+0.017(21)	3.984(21)	1.010(2)	0.979	+0.016(21)
C(4)	4.085(22)	1.010(2)	0.979	-0.084(22)	4.084(22)	1.010(2)	0.979	-0.083(22)	4.085(22)	1.010(2)	0.979	-0.084(22)
C(5)	3.857(28)	1.039(2)	0.99	+0.142(28)	3.859(28)	1.039(2)	0.99	+0.141(28)	3.857(28)	1.039(2)	0.99	+0.143(28)
C(6)	3.924(29)	1.039(2)	0.99	+0.076(29)	3.924(29)	1.039(2)	0.99	+0.075(29)	3.924(29)	1.039(2)	0.99	+0.076(29)

Table 10.5 Net atomic charges for multipole refinements from data refined for multipole refinements where no overload correction (top left), no numerical absorption correction (top right), and no overload or numerical absorption correction (bottom) are applied.

It has not escaped attention that the values in the above tables for BCPs and charges of multipole refinements with 'no overload correction' and 'no overload or numerical correction' are essentially the same for both refinements. The refinements have thus been checked to ensure no mistakes were made and the correct data is reported. It was found that κ' parameters are different in CIF but beyond level of significance reported by *XDGEOM*. Multipole parameters are also different in the *xd_lsm.out* files from the two refinements – but not to the degree of significance reported in the CIFs. Therefore, the fact that both refinements return the same results is potentially odd – but not excessively unusual considering the similarities of the multipolar parameters.

10.1.10 Additional plots and tables for Section 3.6.1

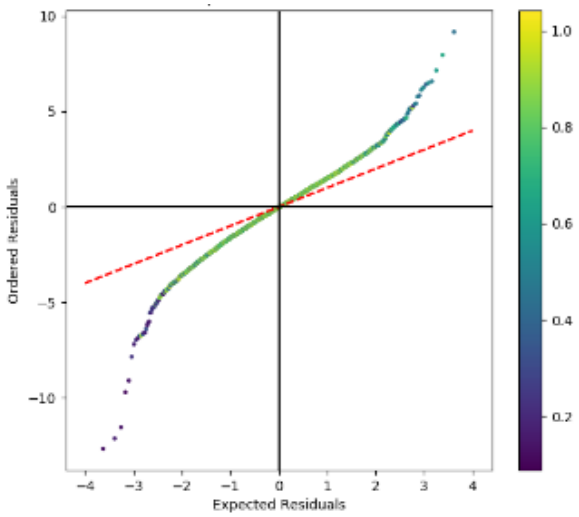


Figure 10.23 Normal probability plot for multipole refinement using data integrated with simple summation algorithm. The colourbar shows the resolution of the reflection in Å^{-1}.

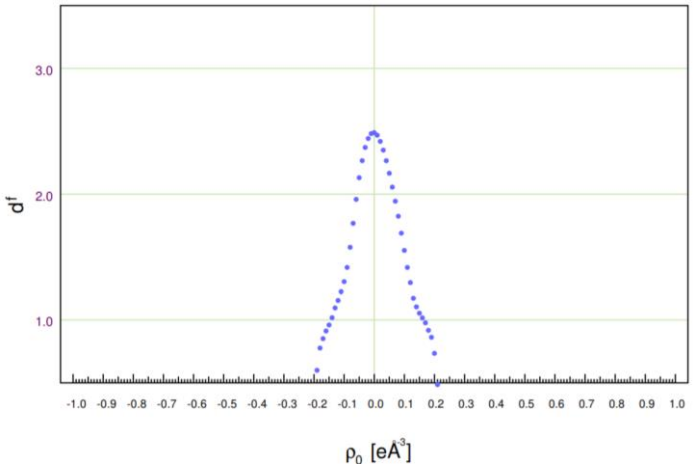


Figure 10.24 Fractal dimensionality plot for multipole refinement using data integrated with simple summation algorithm.

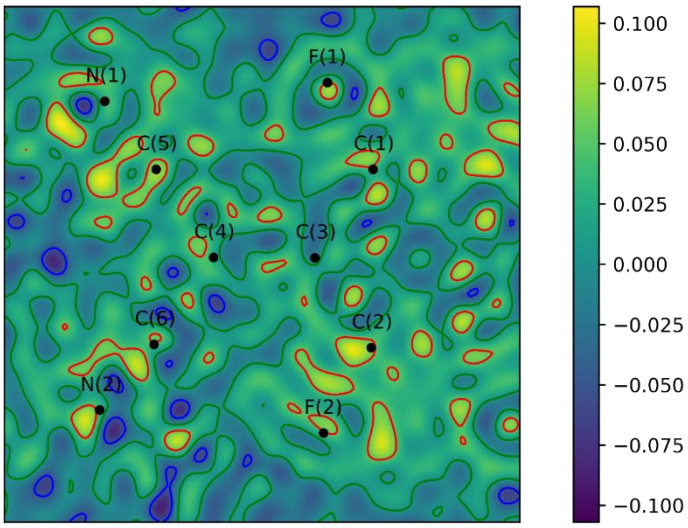


Figure 10.25 Residual electron density plot in the plane of the molecule for multipole refinement with data integrated using simple summation algorithm. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

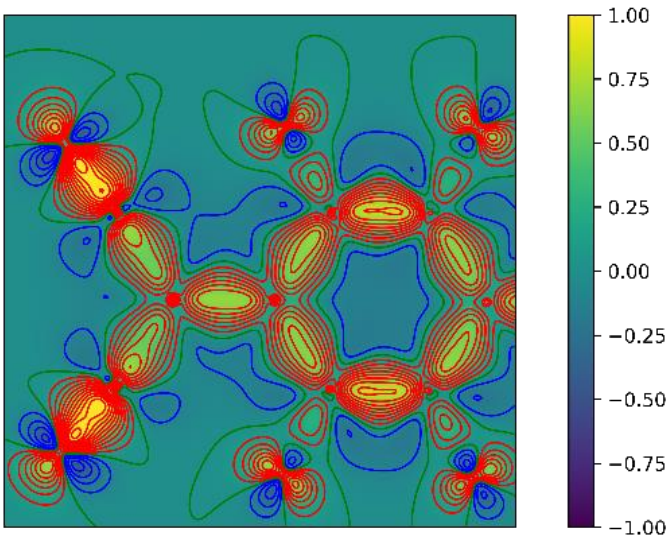


Figure 10.26 Static deformation electron density plot in the plane of the molecule for multipole refinement with data integrated with simple summation algorithm. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.145(15)	-23.380(81)	0.1
F(2)	C(2)	2.099(15)	-21.093(82)	0.08
N(1)	C(5)	3.415(23)	-28.605(190)	0.06
N(2)	C(6)	3.392(22)	-30.032(191)	0.05
C(1)	C(3)	2.011(8)	-16.594(27)	0.17
C(2)	C(3)	2.007(8)	-16.879(28)	0.16
C(1)	C(2*)	2.392(26)	-23.579(65)	0.38
C(3)	C(4)	2.213(9)	-19.997(32)	0.26
C(4)	C(5)	1.920(9)	-13.133(29)	0.08
C(4)	C(6)	1.955(9)	-14.789(29)	0.1

Table 10.6 Properties at BCPs for multipole refinement from data integrated with simple summation algorithm.

Atom	P_{val}	K	κ′	Net atomic charge
F(1)	7.118(11)	0.999	0.922	-0.118(11)
F(2)	7.116(11)	0.999	0.922	-0.116(11)
N(1)	5.100(25)	0.999(2)	1.019	-0.099(25)
N(2)	5.045(26)	0.999(2)	1.019	-0.045(26)
C(1)	3.886(20)	1.019(2)	0.971	+0.113(20)
C(2)	3.873(20)	1.019(2)	0.971	+0.127(20)
C(3)	3.993(22)	1.011(2)	0.961	+0.006(22)
C(4)	4.092(23)	1.011(2)	0.961	-0.091(23)
C(5)	3.840(28)	1.041(2)	0.979	+0.159(28)
C(6)	3.936(29)	1.041(2)	0.979	+0.063(29)

Table 10.7 Net atomic charges for multipole refinement from data integrated with simple summation algorithm.

10.1.11 Additional plots and tables for Section 3.6.2

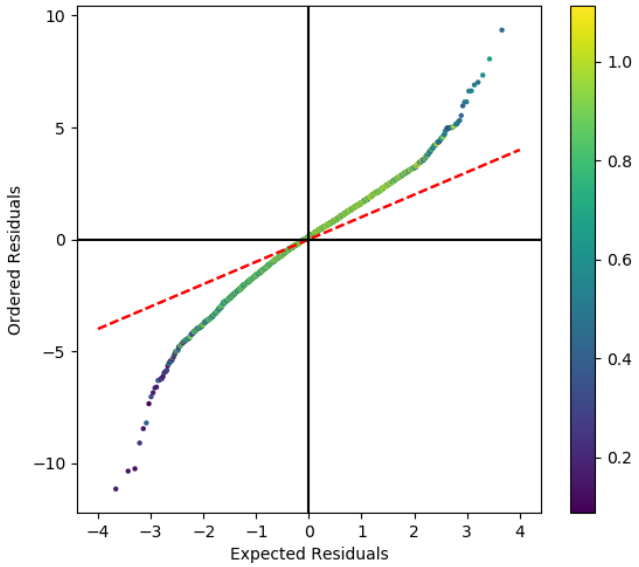


Figure 10.27 Normal probability plot for multipole refinement with 0.45 Å^{-1} resolution cut-off applied. The colourbar shows the resolution of the reflection in Å^{-1}.

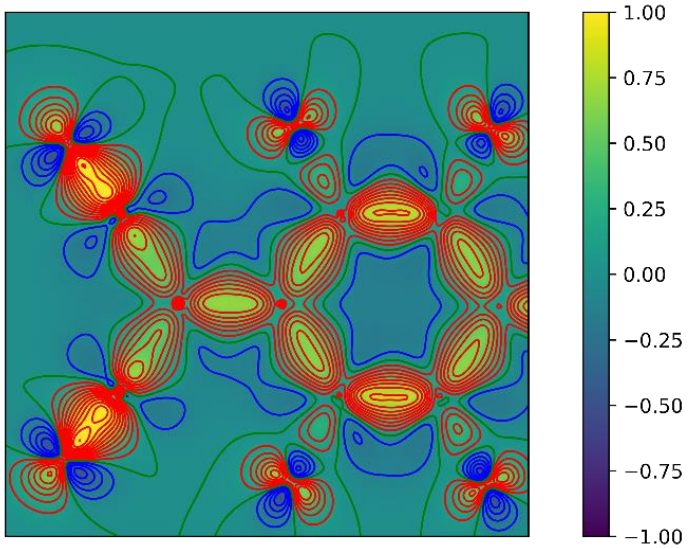


Figure 10.28 Static deformation electron density plot in the plane of the molecule for multipole refinement with 0.45 Å^{-1} resolution cut-off applied. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

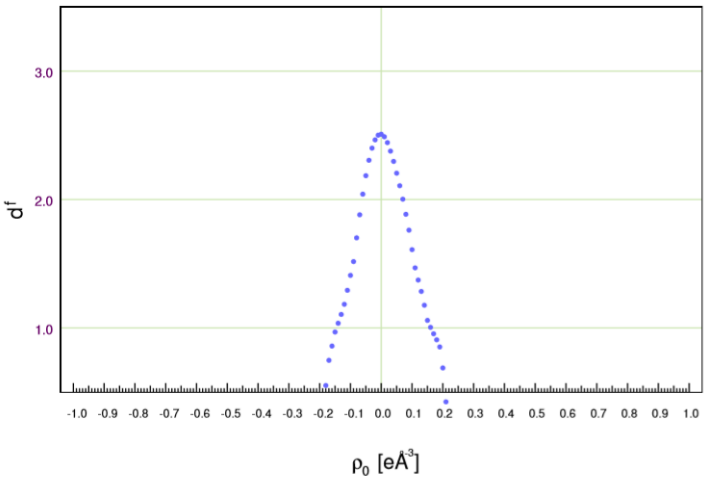


Figure 10.29 Fractal dimensionality plot for 0.45 Å resolution limit refinement.

Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.131(14)	-23.654(82)	0.1
F(2)	C(2)	2.094(14)	-21.971(83)	0.08
N(1)	C(5)	3.390(24)	-28.414(191)	0.06
N(2)	C(6)	3.365(24)	-28.593(196)	0.06
C(1)	C(3)	2.008(8)	-16.709(29)	0.19
C(2)	C(3)	2.009(8)	-17.313(29)	0.16
C(1)	C(2*)	2.393(26)	-23.825(65)	0.39
C(3)	C(4)	2.208(9)	-20.134(35)	0.27
C(4)	C(5)	1.956(9)	-14.962(31)	0.11
C(4)	C(6)	1.911(9)	-13.102(31)	0.08

Table 10.8 Properties at BCPs for multipole refinement from data refined with 0.45 Å^{-1} resolution cut-off applied.

Atom	P_{val}	K	κ'	Net atomic charge
F(1)	7.090(11)	0.998	0.96	-0.089(11)
F(2)	7.087(10)	0.998	0.96	-0.087(10)
N(1)	5.108(24)	0.995(1)	1.01	-0.108(24)
N(2)	5.063(25)	0.995(1)	1.01	-0.063(25)
C(1)	3.917(19)	1.013(2)	0.985	+0.083(19)
C(2)	3.890(18)	1.013(2)	0.985	+0.109(18)
C(3)	3.977(21)	1.008(2)	0.979	+0.023(21)
C(4)	4.078(21)	1.008(2)	0.979	-0.078(21)
C(5)	3.865(27)	1.037(2)	0.989	+0.135(27)
C(6)	3.925(28)	1.037(2)	0.989	+0.074(28)

Table 10.9 Net atomic charges for multipole refinement from data refined with 0.45 Å^{-1} resolution cut-off applied.

10.1.12 Additional plots and tables for Section 3.6.3

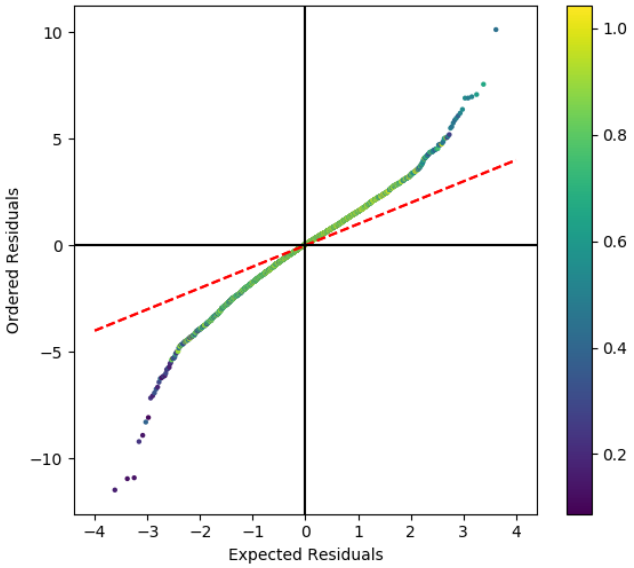


Figure 10.30 Normal probability plot for multipole refinement with phosphor efficiency of 70.5%. The colourbar shows the resolution of the reflection in Å^{-1}.

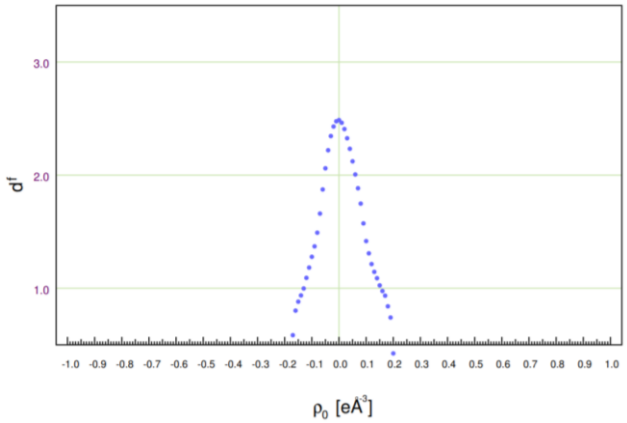


Figure 10.31 Fractal dimensionality plot for multipole refinement with phosphor efficiency of 70.5%.

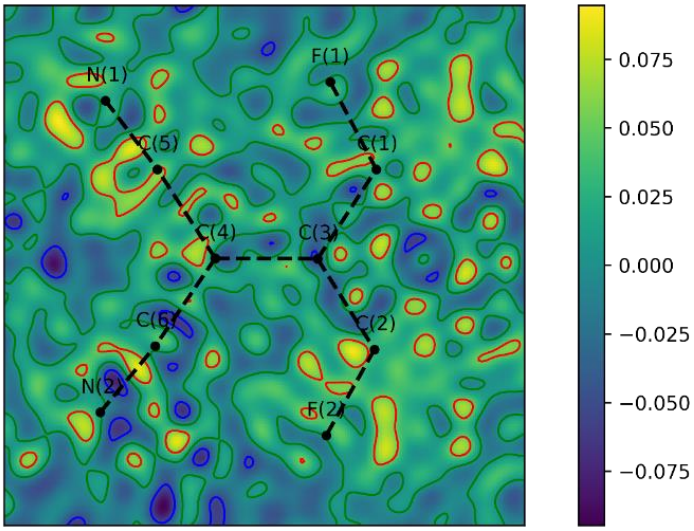


Figure 10.32 Residual electron density plot in the plane of the molecule for multipole refinement with phosphor efficiency of 70.5%. Contour values are drawn at levels of ± 0.05 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours.

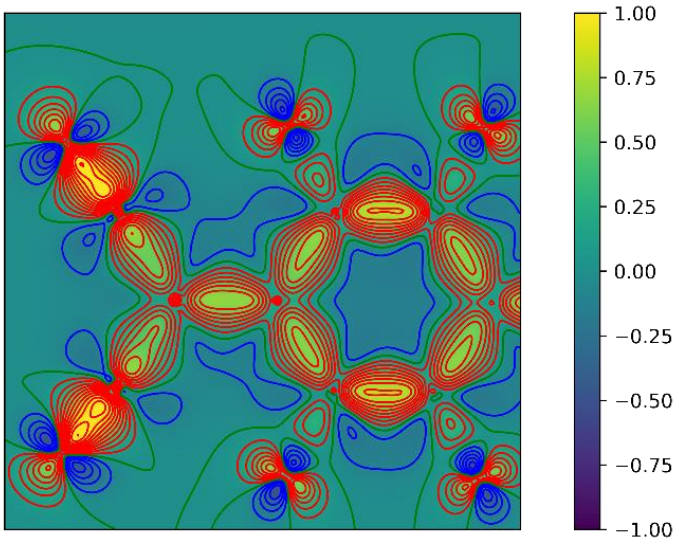


Figure 10.33 Static deformation electron density plot in the plane of the molecule for multipole refinement with phosphor efficiency of 70.5%. Contour values are drawn at levels of ± 0.1 eÅ^{-3}, where red indicates positive contours and blue indicates negative contours. The colourbar is cut-off at ± 1 eÅ^{-3} in order to better view variation of values across the plot.

Atom 1	Atom 2	ρ (eÅ^{-3})	$\nabla^2\rho$ (eÅ^{-5})	ε
F(1)	C(1)	2.137(15)	-23.429(85)	0.1
F(2)	C(2)	2.097(15)	-21.787(86)	0.08
N(1)	C(5)	3.397(25)	-28.943(199)	0.06
N(2)	C(6)	3.372(24)	-29.539(201)	0.06
C(1)	C(3)	2.012(8)	-16.669(30)	0.18
C(2)	C(3)	2.010(9)	-17.112(30)	0.16
C(1)	C(2*)	2.391(26)	-23.639(67)	0.39
C(3)	C(4)	2.211(10)	-20.127(35)	0.27
C(4)	C(5)	1.955(9)	-14.800(31)	0.11
C(4)	C(6)	1.913(9)	-13.011(32)	0.08

Table 10.10 Properties at BCPs for multipole refinements from data refined with phosphor efficiency of 70.5%.

Atom	P_{val}	κ	κ’	Net atomic charge
F(1)	7.103(11)	0.999	0.952	-0.102(11)
F(2)	7.098(11)	0.999	0.952	-0.098(11)
N(1)	5.109(25)	0.996(2)	1.009	-0.108(25)
N(2)	5.060(26)	0.996(2)	1.009	-0.060(26)
C(1)	3.905(19)	1.016(2)	0.984	+0.094(19)
C(2)	3.882(19)	1.016(2)	0.984	+0.118(19)
C(3)	3.983(21)	1.010(2)	0.979	+0.017(21)
C(4)	4.085(22)	1.010(2)	0.979	-0.085(22)
C(5)	3.853(28)	1.040(2)	0.989	+0.146(28)
C(6)	3.922(29)	1.040(2)	0.989	+0.078(29)

Table 10.11 Net atomic charges for multipole refinements from data refined with phosphor efficiency of 70.5%.

10.2 Additional information for Chapter 4

10.2.1 Data collection strategy for F_4TCNQ polymorph II.

Run	Scan	2theta(°)	Omega(°)	Phi(°)	Scan length (s)	Scan width (°)	# of frames
1	Omega	54.26	226.82	200.00	40	0.5	390
2	Omega	54.26	226.82	40.00	40	0.5	390
3	Omega	39.26	211.82	180.00	16	0.5	390
4	Omega	54.26	226.82	160.00	40	0.5	390
5	Omega	56.26	226.82	320.00	40	0.5	390
6	Omega	56.26	226.82	280.00	40	0.5	390
7	Omega	39.26	211.82	270.00	16	0.5	390
8	Omega	56.26	226.82	80.00	40	0.5	390
9	Omega	39.26	211.82	90.00	16	0.5	390
10	Omega	56.26	226.82	0.00	40	0.5	390
11	Omega	35.75	208.31	320.00	10	0.5	390
12	Omega	56.26	226.82	240.00	40	0.5	390
13	Omega	56.26	226.82	120.00	40	0.5	390
14	Omega	35.75	208.31	200.00	10	0.5	390
15	Phi	54.26	61.82	335.49	40	0.5	466
16*	Phi	0.00	0.00	0	1	1	180
17	Phi	35.00	35.00	0	2	1.0	360

Table 10.12 Data collection strategy for F_4TCNQ polymorph II.

The detector distance was 55.00 mm. Run 16 was used only for the correction of overloaded reflections.

10.3 Additional information for Chapter 5

10.3.1 Data collection strategy for L-ascorbic acid

Run	Scan	2theta(°)	Omega(°)	Phi(°)	Scan length (s)	Scan width	# of frames
1	Phi	0	0	0	2	0.5	400
2	Phi	0	0	0	1	0.3	1200
3	Phi	0	180	0	1.5	0.3	1200
4	Omega	0	178.00	0	4	0.3	613
5	Omega	0	178.00	120	6	0.3	613
6	Omega	44.78	222.19	174.4	25	0.30	618
7	Omega	44.9	222.31	320	25	0.3	618
8	Omega	45.55	222,96	213.03	25	0.3	618
9	Omega	45.19	222.60	284.24	25	0.3	618
10	Omega	45.47	222.88	119.3	25	0.3	618
11	Omega	0	238.19	72	8	0.5	118
12	Omega	45.51	222.92	26.86	25	0.3	618
13	Omega	44.9	222.31	80	25	0.3	618
14	Omega	45.88	222.29	341.81	25	0.3	618
15	Omega	44.71	222.12	160.11	25	0.3	618
16	Omega	44.9	222.31	40	25	0.3	618
17	Omega	45.62	222.03	132.24	20	0.3	618
18	Omega	44.06	221.47	303.08	20	0.3	618
19	Omega	43.99	221.40	95.08	25	0.3	618
20	Omega	44.9	222.31	0	30	0.3	618
21	Omega	44.9	222.31	200	30	0.3	618
22	Omega	14.9	192.31	90	10	0.3	618
23	Omega	-14.9	162.92	90	13	0.3	618
24*	Phi	44.9	222.08	72.00	3	0.3	113
25	Phi	44.9	222.08	72.00	3	0.3	113

*attenuated

Table 10.13 Data collection strategy for L-ascorbic acid. Runs 1, 2, 24 and 25 were used only to correct overloaded reflections (as 'fast scans' in SADABS).

10.3.2 White radiation contamination

White radiation contamination was found in L-ascorbic acid data. The reflections that were removed were those most affected by the contamination. These are listed below:

- 0 0 7
- 2 0 3
- 0 0 7
- 0 7 7
- 0 -7 -7
- 0 8 8
- 0 -8 -8
- 0 9 9
- 0 -9 -9
- -8 0 20

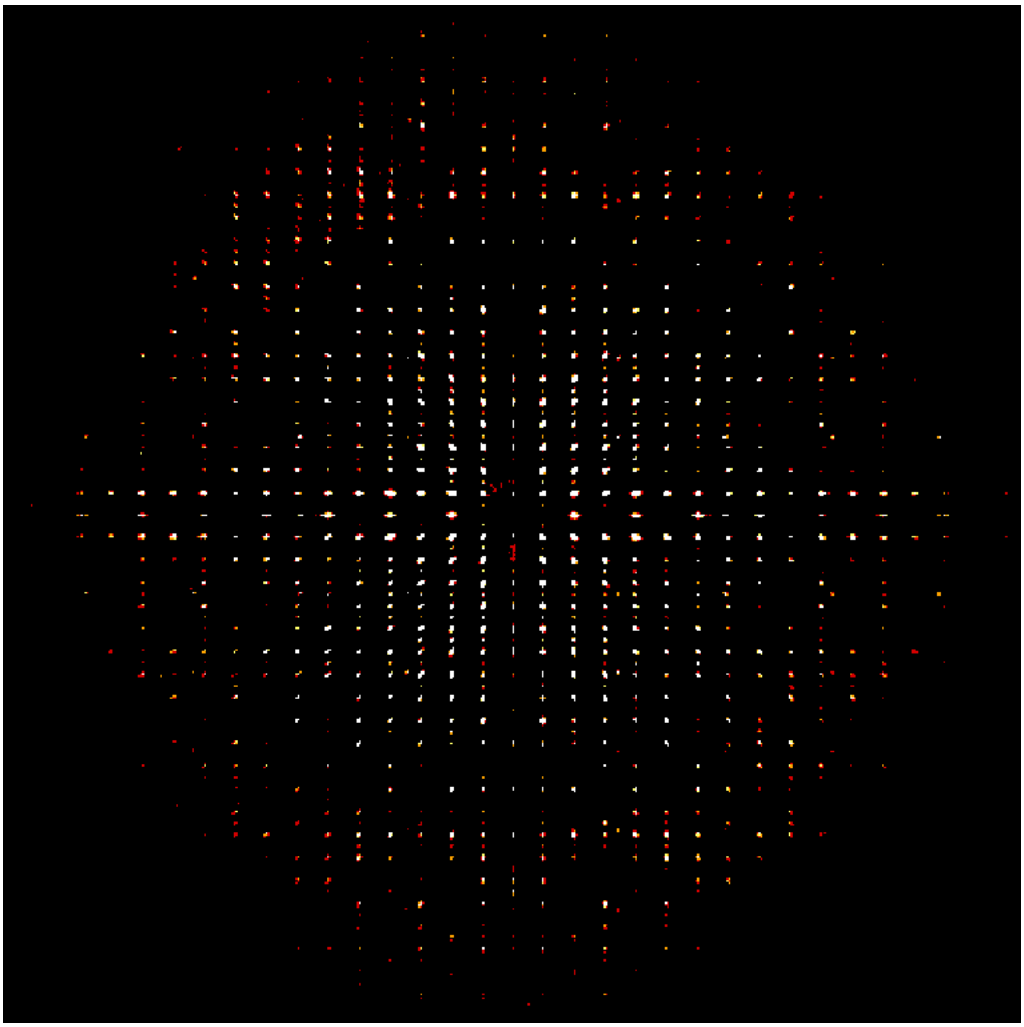
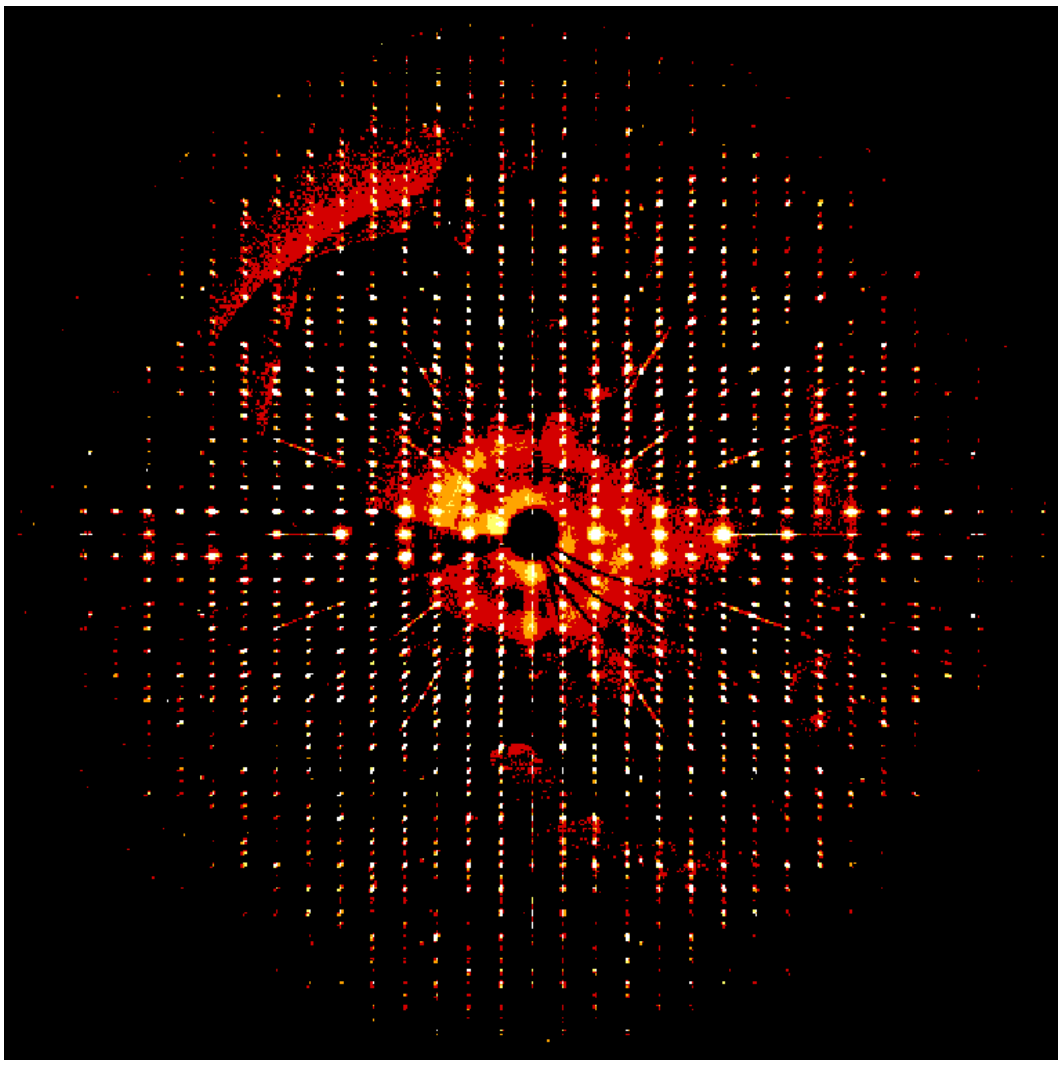


Figure 10.34 Precession image calculated with all experimental data. No contamination from white radiation appears to be present.



10.1 Precession Image without fast scans included. In this image the white radiation contamination is visible.

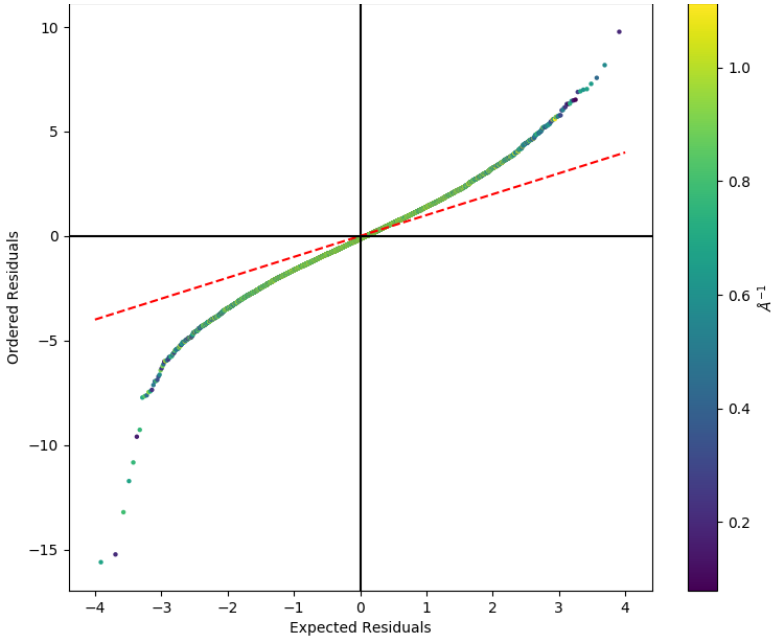


Figure 10.35 Normal probability plot for statistically weighted multipole model of L-ascorbic acid, using all data.

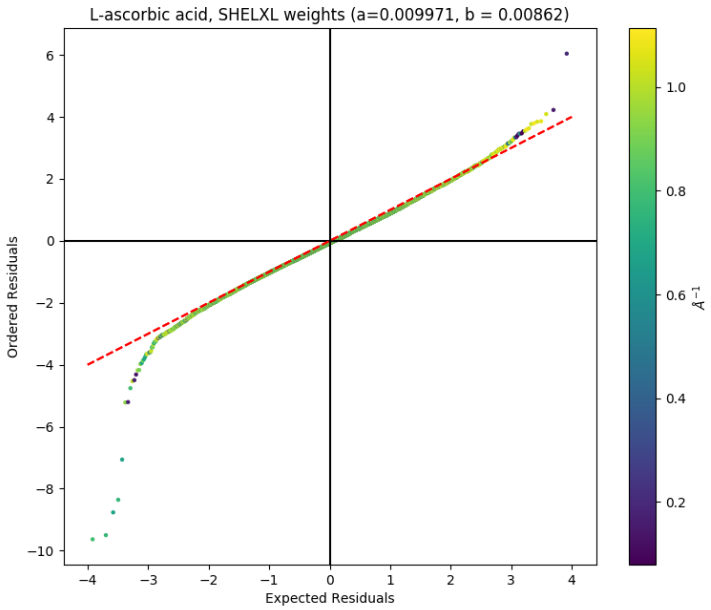


Figure 10.36 Normal probability plot for SHELXL weighted refinement containing reflections contaminated with white radiation. Optimised values for SHELXL weighting scheme were calculated as a = 0.00997, b = 0.00862.

10.3.3 Multipolar refinement strategy list for L-ascorbic acid

Key: Monopoles, dipoles, quadrupoles, octupoles and hexadecapoles denoted by the initial (M, D, Q, O, H). K = Kappa, K' = Kappa 2.

Setup refinement - as *XDINI* does not carry over all Uijs from *SHELX* ins file, therefore these values must be refined in *XD*. Hydrogen Uijs were calculated using SHADE server,

1. Scale only, statistical weights
2. Non-hydrogen xyz and Uij
3. Non-hydrogen xyz (sinθ 0.7 to 1.0)
4. Non-hydrogen xyz, Uij (sinθ 0.7 to 1.0)
5. Non-hydrogen scale only (sinθ 0.7 to 1.0)
6. Non-hydrogen xyz, Uij (sinθ 0.7 to 1.0)

Multipolar refinement – I/σ < 3 removed, statistical weights. The maximum multipole level refined in all cases for hydrogen atoms was one bond directed dipole, therefore where Q, O, H are concerned they refer to only non-hydrogen atoms. Hydrogen positions and Uijs are not refined unless otherwise stated. Non-hydrogen κ are fixed at 1.2 and also not refined. H bonds are reset to expect values in all iterations of the refinement unless otherwise stated.

1. Non-hydrogen monopoles
2. All monopoles
3. D Q O H
4. M, D, Q, O, H
5. Non-hydrogen Uij
6. M, D, Q, O, H
7. Non-hydrogen Uij; M, D, Q, O, H
8. Non-hydrogen xyz
9. Non-hydrogen xyz; M, D, Q, O, H
10. Non-hydrogen xyz, Uij; M, D, Q, O, H
11. κ
12. M
13. κ, M
14. Non-hydrogen xyz, Uij; M, D, Q, O, H
15. κ
16. Non-hydrogen xyz, Uij; M, D, Q, O, H
17. Hydrogen xyz (sinθ/λ cut-off 0 – 0.5 $Å^{-1}$) no reset hydrogen bond
18. Non-hydrogen xyz, Uij; M, D, Q, O, H
19. κ, Non-hydrogen xyz, Uij; M, D, Q, O, H
20. κ, D, Q, O, H
21. κ'
22. D, Q, O, H
23. κ, Non-hydrogen xyz, Uij; M, D, Q, O, H

SHADE server input updated hydrogen Uij

1. Hydrogen xyz (sinθ/λ cut-off 0 – 0.5 $Å^{-1}$) no reset hydrogen bond
2. Non-hydrogen xyz, Uij; D, Q, O, H
3. κ, Non-hydrogen xyz, Uij; D, Q, O, H
4. κ, Non-hydrogen xyz, Uij; M, D, Q, O, H
5. κ'
6. κ, Non-hydrogen xyz, Uij; M, D, Q, O, H

10.3.4 Additional plots for Chapter 5

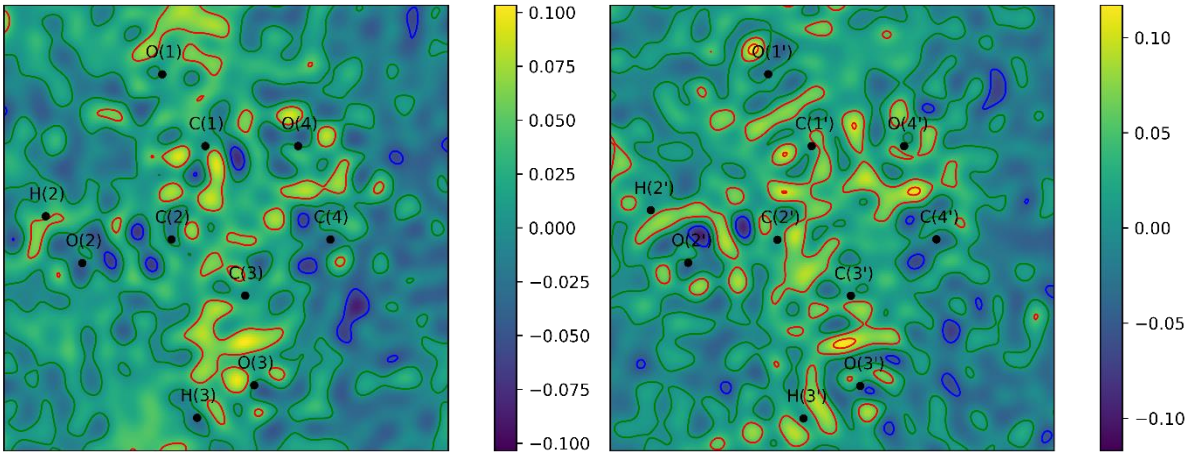


Figure 10.37 Residual electron density plots in plane of ring for molecule 1 and 2 for a statistical weighted model of L-ascorbic acid. Contours are drawn at ± 0.05 eÅ-3 levels, where red indicates positive contours and blue indicates negative contours. The zero contour is present in green.

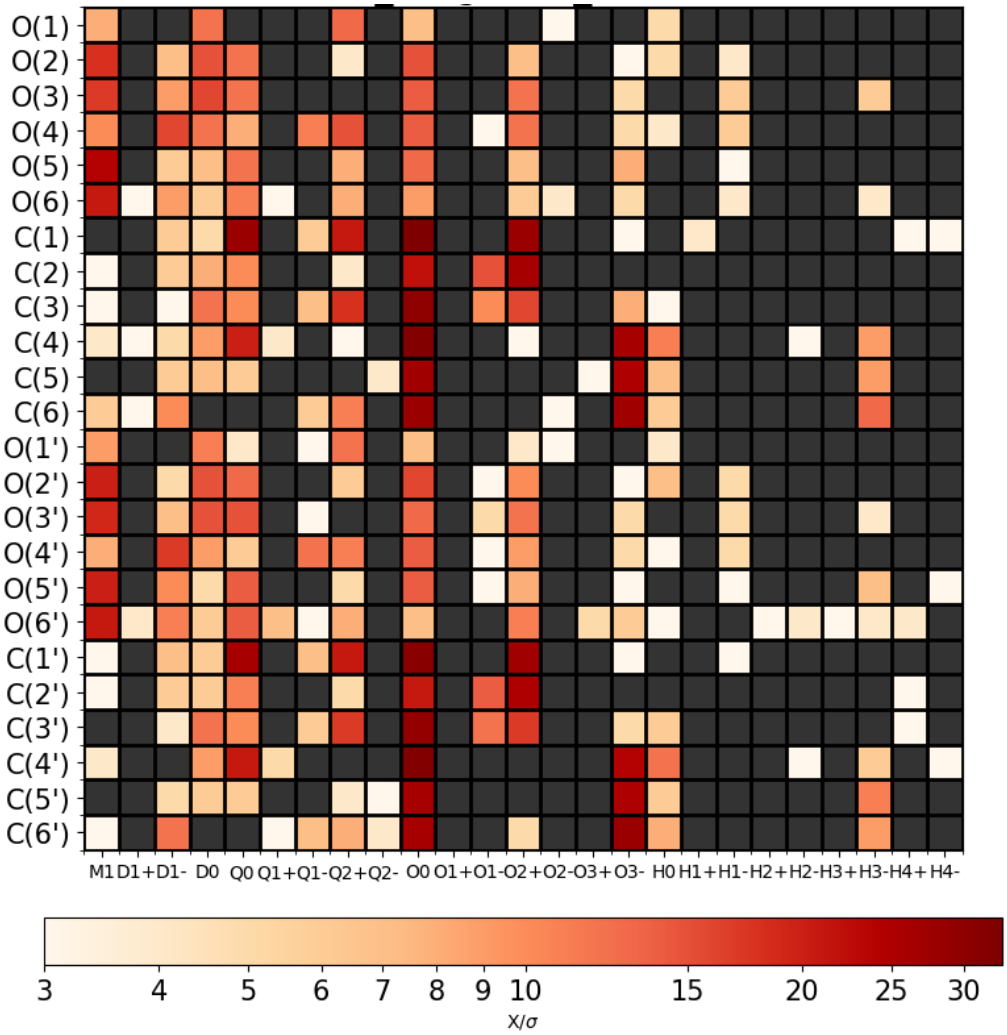


Figure 10.38 Significance of L-ascorbic acid multipolar parameters, where those with X/σ < 3 are coloured dark grey.

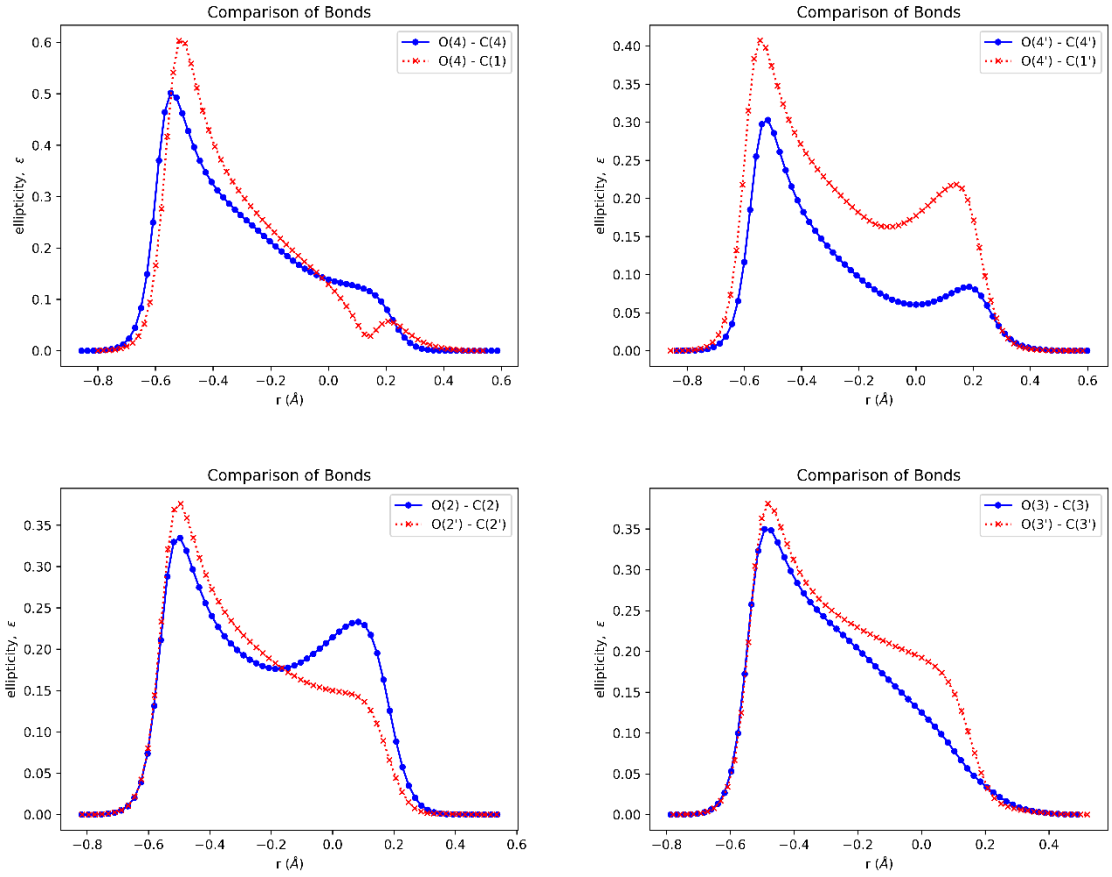
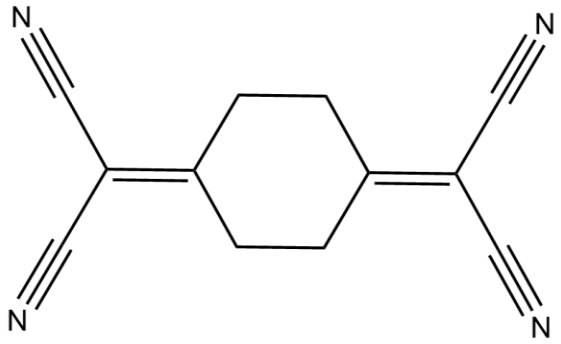


Figure 10.39 Comparison of the ellipticity along the BCP of bonds within molecules 1 and 2 of L-ascorbic acid.

10.4 Additional Information for Chapter 6

10.4.1 Synthesis of 1,4-bis(dicyanomethylene)cyclohexane



Scheme 10.1 1,4-bis(dicyanomethylene)cyclohexane (DCMC).

1,4-bis(dicyanomethylene)cyclohexane was synthesised using the method described by Acker and Hertler.[189] A mixture of 1,4-cyclohexanedione (1.0 g, 8.9 mmol) and malononitrile (1.1 ml, 18.0 mmol) were melted together at 80 °C for 5 minutes, then treated with water (5 ml) containing a catalytic amount of alanine (10 mg, 0.1 mmol). Heating was continued with stirring for about 10 minutes until crystals began to form. The mixture was the cooled to room temperature and the resultant solid was collected by filtration, washed with water (10 ml) and Et_2O (10 ml) to afford the title compound as a light brown solid (1.3 g, 70 %); δ_H (400 MHz, $CDCl_3$) 2.99 (8H, s, (CH2)4); δ_C (101 MHz, $CDCl_3$) 175.42, 110.48 (CN), 86.86, 31.39 $(CH_2)_4$); all data agree with those reported in the literature.[240]

10.4.2 Data collection strategy for TCNQ and DCMC

Run	Scan	2theta(°)	Omega(°)	Phi(°)	Scan length (s)	Scan width	# of frames
1	Omega	54.26	226.82	320.00	40	0.5	390
2	Omega	54.26	226.82	80.00	40	0.5	390
3	Omega	39.26	211.82	180.00	20	0.5	390
4	Omega	54.26	226.82	240.00	40	0.5	390
5	Omega	39.26	211.82	90.00	20	0.5	390
6	Omega	54.26	226.82	160.00	40	0.5	390
7	Omega	54.26	226.82	320.00	40	0.5	390
8	Omega	39.26	211.82	270.00	20	0.5	390
9	Omega	53.36	225.92	18.15	37.4	0.5	390
10	Omega	53.36	225.92	309.00	37.4	0.5	390
11	Omega	39.26	211.82	0.0	20	0.5	390
12	Omega	54.26	226.82	240.00	40	0.5	390

Table 10.14 Data collection strategy for TCNQ.

Run	Scan	2theta(°)	Omega(°)	Phi(°)	Scan length (s)	Scan width	# of frames
1	Omega	0.00	0.00	0.00	1	1.0	180
2	Omega	42.40	214.34	120.00	40	0.5	393
3	Omega	41.51	213.44	301.58	32	0.5	393
4	Omega	42.40	214.34	40.00	40	0.5	393
5	Omega	41.51	213.44	139.95	32	0.5	393
6	Omega	35.37	207.30	320.00	10	0.5	393
7	Omega	42.40	214.34	240.00	40	0.5	393
8	Omega	42.40	214.34	200.00	40	0.5	393
9	Omega	42.40	214.34	0.00	40	0.5	75

Table 10.15 Data collection strategy for DCMC.

10.4.3 Multipolar refinement parameter list for TCNQ/DCMC

Key: Monopoles, dipoles, quadrupoles, octupoles and hexadecapoles denoted by the initial (M, D, Q, O, H). K = Kappa, K' = Kappa 2.

Setup refinement - as *XDINI* does not carry over all Uijs from *SHELX* ins file, therefore these values must be refined in *XD*. Hydrogen Uijs were calculated using SHADE server,

8. Scale only, statistical weights
9. Non-hydrogen xyz and Uij
10. Non-hydrogen xyz (sinθ 0.7 to 1.0)
11. Non-hydrogen xyz, Uij (sinθ 0.7 to 1.0)
12. Non-hydrogen scale only (sinθ 0.7 to 1.0)
13. Non-hydrogen xyz, Uij (sinθ 0.7 to 1.0)

Multipolar refinement – I/σ < 3 removed, statistical weights. The maximum multipole level refined in all cases for hydrogen atoms was the dipolar, therefore where Q, O, H are concerned they refer to only non-hydrogen atoms. Hydrogen positions and Uijs are not refined unless otherwise stated. Non-hydrogen κ are fixed at 1.2 and also not refined. H bonds are reset to expect values in all iterations of the refinement unless otherwise stated.

24. Non-hydrogen M
25. All M
26. D, Q, O, H
27. M, D, Q, O, H
28. Uij
29. M, D, Q, O, H
30. Uij, M, D, Q, O, H
31. xyz
32. xyz, M, D, Q, O, H
33. xyz, Uij, M, D, Q, O, H
34. κ
35. M
36. K, M
37. xyz, Uij, M, D, Q, O, H
38. κ
39. xyz, Uij, M, D, Q, O, H
40. κ, M, D, Q, O, H
41. xyz, Uij, M, D, Q, O, H
42. No Reset H bond, h xyz, sin 0 – 0.5
43. xyz, Uij, D, Q, O, H
44. κ'
45. D, Q, O, H
46. κ, xyz, Uij, M, D, Q, O, H
47. κ '
48. D, Q, O, H
49. xyz, Uij, M, D, Q, O, H
50. κ
51. κ, xyz, Uij, M, D, Q, O, H

H atom Uij values recalculated from SHADE server

52. No Reset H bond, h xyz, sin 0 – 0.5
53. xyz, Uij, D, Q, O, H
54. κ, xyz, Uij, D, Q, O, H
55. κ, xyz, Uij, M, D, Q, O, H
56. κ'
57. κ, xyz, Uij, M, D, Q, O, H

10.4.4 Additional plots for Chapter 6

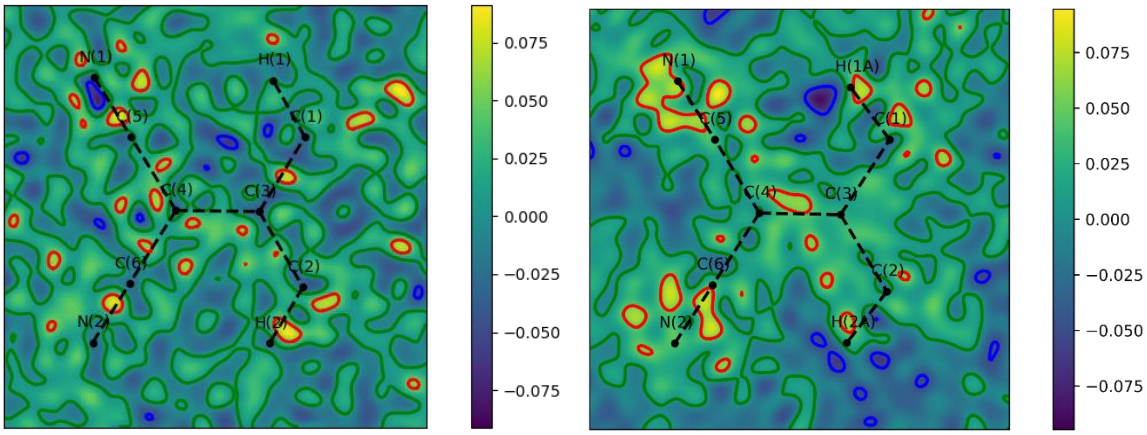


Figure 10.40 Residual electron density plot of TCNQ (left) and DCMC (right) for statistically weighted molecules. Contours are drawn at ± 0.05 eÅ-3 levels, where red indicates positive contours and blue indicates negative contours. The zero contour is present in green.

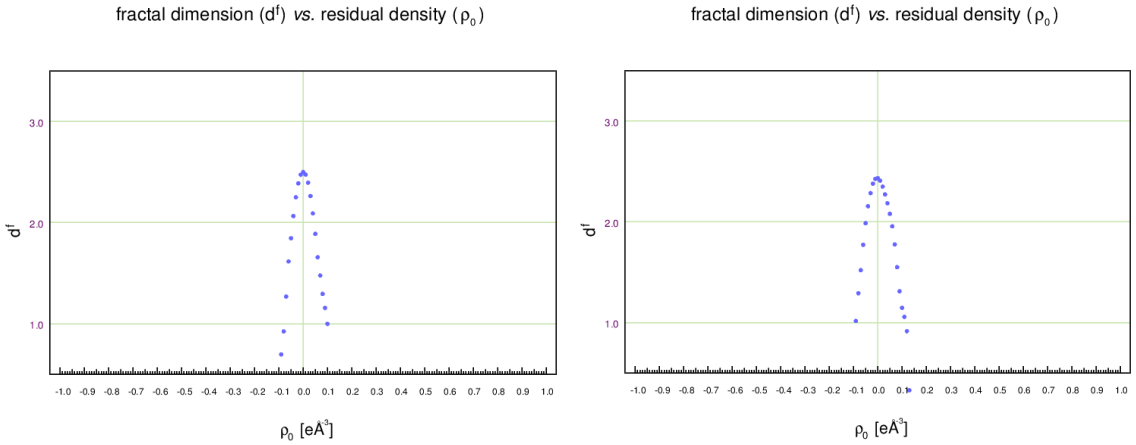


Figure 10.41 Fractal dimensionality plot for statistically weighted model of TCNQ (left) and DCMC (right).

10.5 Byte offset compression

To unpack the image data within a *.cbf* frame, the start of the image binary (indicated by a binary marker, "\0xc\x1a\x04\xd5") must be located.[201] Starting with a Base value of 0, the first byte is read and assigned to variable Δ. 1 byte of binary can only store a signed value of up to ± 127. If Δ is -127 ≥ Δ ≥ 127 then Δ is taken as the actual value added to Base value. The new Base value is added to the image matrix (a list of all the decompressed values). The process begins again with the next byte being assigned as Δ.

However, a Δ of -128 or +128 means the actual value is too large to be stored in 1 byte of binary, and so Δ is discarded. The next two bytes are unpacked from the binary as one 16-bit value, which is set as Δ. The maximum signed value that can be stored in 2 bytes of binary is ± 32767. If Δ is between these values then, as before, Δ is added to Base. The new Base value is added to the image matrix and the algorithm begins again, with the next byte being assigned as Δ.

If the 16-bit value of Δ is ± 32768, then the actual value is too large to be stored in 2 bytes of binary and the next four bytes are assigned to Δ as one 32-bit value. Similarly, if this value is between ± -2147483647 (the maximum range of signed values that can be stored using 4 bytes), Δ is added to Base, the new Base value is added to the image matrix, and the process begins again. If Δ is 2147483648, then the actual value is too large to be stored using 4 bytes. The next eight bytes are assigned to Δ as one 64-bit value. This value is added to Base, which is then added to the image matrix. It is assumed no value higher than that can be stored using 8 bytes will be present within the diffraction image.

10.6 Sfrm diffraction format

The *APEX2* integration program, created by Bruker Axs Ltd, requires diffraction frames to be in *sfrm* format. This consists of files with an ASCII header, followed by the binary encoded image, overflow and underflow tables.

The header information is stored in blocks of 512 bytes. If the block is not large enough to meet these criteria, padding is added to the end of the header to make it the requisite length. There are always a multiple of 5 blocks in the header. Each line in the header is 80 bytes long - containing an 8-byte identifier for the header item, followed by 72 bytes of information. The header information contains all the metadata about the experiment and measurement of the frame.

The intensity values of an image are encoded as either 1, 2- or 4-byte integers, as denoted in the file's header. If an image value is larger than can be stored in the binary, the actual value will be stored within an overflow table. The integers are only stored as positive values, any negative values are stored in an underflow table. In an image stored as 1-byte integers, any value of 255 within the frame indicates that the actual intensity value is either greater than or equal to a value of 255 and should be looked up within the first overflow table, in which all intensity values are stored as 2-byte integers. Any value too large for this overflow table will be stored in the second overflow table, in which values are stored as 4-byte integers. Values of 65535 in the first overflow table indicate that the actual value is either 65535 or greater, and the actual value is stored in the second overflow table. Likewise, any intensity value below 0 can be stored in an underflow table.

The overflow and underflow tables are found at the end of the *.sfrm* file, after the image binary. Like the header, the overflow and underflow tables are padded, if they are not of a multiple of 16 bytes in length.

10.7 raw file

The .raw file is a fixed format ASCII or binary file which contains details about each reflection measured in the integration process. A screenshot of a section of the file is shown in Figure 10.42. This file is a standard in the *APEX* programs, but other integration software, for instance *CrysAlisPro*[144], can produce an *APEX* style *.raw* file.

Recorded details include;[200]

- Integrated intensities, which have had corrections for Lorentz effect and polarization applied. These values are also background subtracted and normalised (1 min per degree).
- Standard deviation of intensity based on counting statistics.
- Observed and calculated pixel coordinates of the reflection scaled to those of a 512x512 detector.
- Direction cosines of negative incident and diffracted beam, relative to the reciprocal unit cell axes.
- Total raw peak count (in photons). No corrections have been applied to this value.
- Average background value which is documented per pixel in photons x 1000.
- Unique reflection code, a number given to every reflection based on its *hkl*.
- Correction factors applied to the integrated intensity. The Lorentz, polarization, air absorption, face plate absorption and phosphor absorption corrections are reported as a product. The Lorentz correction applied is also reported on its own.
- Resolution of the reflection. This is recorded in the file as $\frac{\sin\theta}{\lambda} \times 10000$.

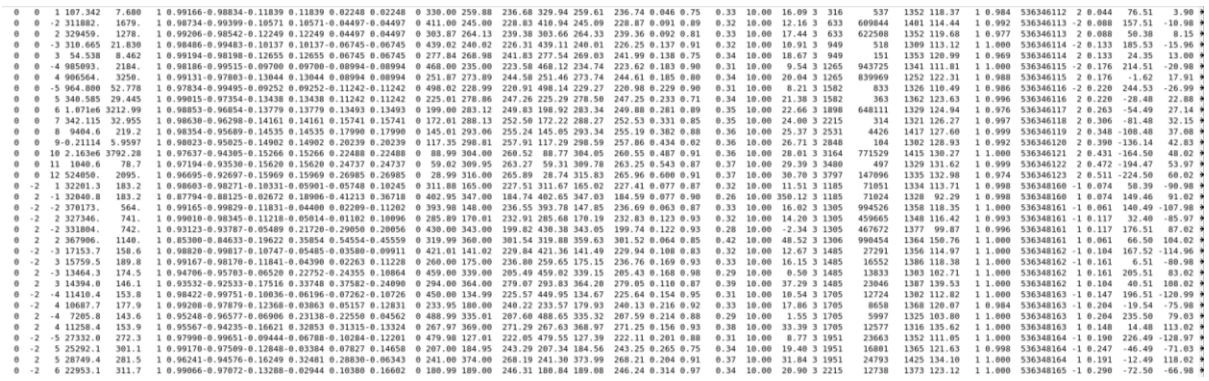


Figure 10.42 Screenshot of part of raw file.

Chapter 11. References

1 T. S. Koritsanszky and P. Coppens, *Chem. Rev.*, 2001, **101**, 1583–627.
2 G. Ferry, *Nature*, 2014, **505**, 609–611.
3 E. F. Garman, *Science*, 2014, **343**, 1102–1108.
4 N. Jones, *Nat. News*, 2014, **505**, 602.
5 C. Giacovazzo, H. L. Monaco, D. Viterbo, F. Scordari, G. Gilli, G. Zanotti and M. Catti, *Fundamentals of Crystallography*, Oxford University Press, 1992.
6 A. H. Compton, *Phys Rev*, 1923, **21**, 483–502.
7 W. H. Bragg and W. L. Bragg, *Proc R Soc Lond A*, 1913, **88**, 428–438.
8 C. Hammond, *The Basics of Crystallography and Diffraction*, Oxford Science Publications, 1997.
9 W. Kabsch, *J. Appl. Crystallogr.*, 1988, **21**, 67–72.
10 W. Clegg, *J. Appl. Crystallogr.*, 1984, **17**, 334–336.
11 W. Kabsch, *J. Appl. Crystallogr.*, 1993, **26**, 795–800.
12 W. Kabsch, *J. Appl. Crystallogr.*, 1977, **10**, 426–429.
13 W. Cochran, *Acta Crystallogr.*, 1952, **5**, 65–67.
14 O. V. Dolomanov, L. J. Bourhis, R. J. Gildea, J. A. K. Howard and H. Puschmann, *J. Appl. Crystallogr.*, 2009, **42**, 339–341.
15 B. Dittrich and C. F. Matta, *IUCrJ*, 2014, **1**, 457–469.
16 P. Munshi and T. N. Guru Row, *Cryst. Growth Des.*, 2006, **6**, 708–718.
17 V. R. Hathwar, R. Pal and T. N. Guru Row, *Cryst. Growth Des.*, 2010, **10**, 3306–3310.
18 N. K. Hansen and P. Coppens, *Acta Crystallogr. Sect. A*, 1978, **34**, 909–921.
19 P. Macchi, *Crystallogr. Rev.*, 2013, **19**, 58–101.
20 M. Milanesio, R. Bianchi, P. Ugliengo, C. Roetti and D. Viterbo, *J. Mol. Struct. THEOCHEM*, 1997, **419**, 139–154.
21 R. F. W. Bader, *J. Phys. Chem. A*, 2009, **113**, 10391–10396.
22 C. S. Lopez and A. R. de Lera, *Curr. Org. Chem.*, 2011, **15**, 3576–3593.
23 R. F. W. Bader, T. S. Slee, D. Cremer and E. Kraka, *J. Am. Chem. Soc.*, 1983, **105**, 5061–5068.
24 J. Henn and K. Meindl, *Acta Crystallogr. Sect. A*, 2014, **70**, 248–256.
25 D. Borek, W. Minor and Z. Otwinowski, *Acta Crystallogr. Sect. D*, 2003, **59**, 2031–2038.
26 K. Diederichs, *Acta Crystallogr. Sect. D*, 2010, **66**, 733–740.
27 H. Wolf, M. R. V. Jørgensen, Y.-S. Chen, R. Herbst-Irmer and D. Stalke, *Acta Crystallogr. Sect. B*, 2015, **71**, 10–19.
28 Bruker AXS Inc., *SADABS*, Bruker AXS Inc., Madison, Wisconsin, USA, 2001.
29 M. S. Weiss and R. Hilgenfeld, *J. Appl. Crystallogr.*, 1997, **30**, 203–205.
30 M. S. Weiss, *J. Appl. Crystallogr.*, 2001, **34**, 130–135.
31 M. S. Weiss, H. J. Metzner and R. Hilgenfeld, 1998, **423**, 291–296.
32 J. Henn and A. Schönleber, *Acta Crystallogr. Sect. A*, 2013, **69**, 549–558.
33 J. Henn, *Acta Crystallogr. Sect. A*, 2016, **72**, 696–703.
34 A. I. Stash, *DRKplot for XD and SHELX*, Moscow, 2007.
35 A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. Mallinson, T. Richter and T. Koritsanszky, *XD2016*, University at Buffalo, State University of New York, NY, USA; University of Milano, Italy; University of Glasgow, UK; CNRISTM, Milano, Italy; Middle Tennessee State University, TN, USA., 2016.

36 C. B. Hübschle and B. Dittrich, *J. Appl. Crystallogr.*, 2011, **44**, 238–240.
37 J. Henn and K. Meindl, *Acta Crystallogr. Sect. A*, 2008, **64**, 404–418.
38 V. V. Zhurov, E. A. Zhurova and A. A. Pinkerton, *J. Appl. Crystallogr.*, 2008, **41**, 340–349.
39 S. C. Abrahams and E. T. Keve, *Acta Crystallogr. Sect. A*, 1971, **27**, 157–165.
40 J. Henn and K. Meindl, *Int. J. Mater. Chem. Phys.*, 2016, **1**, 417–430.
41 J. Henn and K. Meindl, *Acta Crystallogr. Sect. A*, 2015, **71**, 203–211.
42 J. Henn and K. Meindl, *Acta Crystallogr. Sect. A*, 2014, **70**, 499–513.
43 A. Volkov, P. Macchi, L. J. Farrugia, C. Gatti, P. Mallinson, T. Richter and T. Koritsanszky, *XD2016 Manual*, 2016.
44 F. L. Hirshfeld, *Acta Crystallogr. Sect. A*, 1976, **32**, 239–244.
45 R. E. Rosenfield, K. N. Trueblood and J. D. Dunitz, *Acta Crystallogr. Sect. A*, 1978, **34**, 828–829.
46 S. S. Shapiro and M. B. Wilk, *Biometrika*, 1965, **52**, 591–611.
47 R. Kamiński, S. Domagała, K. N. Jarzembska, A. A. Hoser, W. F. Sanjuan-Szklarz, M. J. Gutmann, A. Makal, M. Malińska, J. M. Bąk and K. Woźniak, *Acta Crystallogr. Sect. A*, 2014, **70**, 72–91.
48 E. D. Stevens and P. Coppens, *Acta Crystallogr. Sect. A*, 1976, **32**, 915–917.
49 R. Bianchi, G. Gervasio and D. Marabello, *Comptes Rendus Chim.*, 2005, **8**, 1392–1399.
50 B. Guillot, L. Viry, R. Guillot, C. Lecomte and C. Jelsch, *J. Appl. Crystallogr.*, 2001, **34**, 214–223.
51 X. Li, G. Wu, Y. A. Abramov, A. V. Volkov and P. Coppens, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12132–12137.
52 B. Guillot, N. Muzet, E. Artacho, C. Lecomte and C. Jelsch, *J. Phys. Chem. B*, 2003, **107**, 9109–9121.
53 R. Destro, L. Loconte, L. Lo Presti, P. Roversi and R. Soave, *Acta Crystallogr. Sect. A*, 2004, **60**, 365–370.
54 B. T. M. Willis, in *International Tables for Crystallography*, 2006, vol. B, pp. 400–406.
55 W. Massa, *Crystal Structure Determination*, Springer-Verlag Berlin Heidelberg, 2004.
56 J. Oddershede and S. Larsen, *J. Phys. Chem. A*, 2004, **108**, 1057–1063.
57 B. Niepötter, R. Herbst-Irmer and D. Stalke, *J. Appl. Crystallogr.*, 2015, **48**, 1485–1497.
58 M. I. McMahon, *Top. Curr. Chem.*, 2012, **315**, 69–109.
59 L. Krause, R. Herbst-Irmer, G. M. Sheldrick and D. Stalke, *J. Appl. Crystallogr.*, 2015, **48**, 3–10.
60 R. H. Blessing, *J. Appl. Crystallogr.*, 1997, **30**, 421–426.
61 D. Leusser, in *Electron Density and Chemical Bonding I*, ed. D. Stalke, Springer Berlin Heidelberg, 2012, pp. 101–126.
62 L. Krause, B. Niepötter, C. J. Schürmann, D. Stalke and R. Herbst-Irmer, *IUCrJ*, 2017, **4**, 420–430.
63 H. O. Sørensen and S. Larsen, *J. Appl. Crystallogr.*, 2003, **36**, 931–939.
64 S. M. Gruner, M. W. Tate and E. F. Eikenberry, *Rev. Sci. Instrum.*, 2002, **73**, 2815–2842.
65 A. Martin and A. A. Pinkerton, *Acta Crystallogr. Sect. B*, 1998, **54**, 471–477.
66 D. Waterman and G. Evans, *J. Appl. Crystallogr.*, 2010, **43**, 1356–1371.
67 P. Allé, E. Wenger, S. Dahaoui, D. Schaniel and C. Lecomte, *Phys. Scr.*, 2016, **91**, 063001.
68 A. C. Thompson, E. M. Westbrook, W. M. Lavender and J. C. Nix, *J. Phys. Conf. Ser.*, 2014, **493**, 012019.
69 H. Graafsma, *J. Synchrotron Radiat.*, 2016, **23**, 383–384.
70 Bruker AXS Inc., *Technical Note SC-XRD 12: New Developments in Pixel Array Technology*, 2016.
71 A. G. W. Leslie, *Acta Crystallogr. Sect. D*, 1999, **55**, 1696–1702.
72 Bruker AXS Inc., *SAINT*, Bruker AXS Inc., Madison, Wisconsin, USA, 2012.

73 J. W. Pflugrath, *Acta Crystallogr. Sect. D*, 1999, **55**, 1718–1725.
74 G. C. Ford, *J. Appl. Crystallogr.*, 1974, **7**, 555–564.
75 Bruker AXS Inc., *APEX3*, Bruker AXS Inc., Madison, Wisconsin, USA, 2015.
76 Bruker AXS Inc., *SADABS User Manual (version 2.03)*, 2002.
77 P. Coppens, J. de Meulenaer and H. Tompa, *Acta Crystallogr.*, 1967, **22**, 601–602.
78 W. R. Busing and H. A. Levy, *Acta Crystallogr.*, 1957, **10**, 180–182.
79 J. Henn and K. Meindl, *Acta Crystallogr. Sect. A*, 2010, **66**, 676–684.
80 M. R. V. Jørgensen, H. Svendsen, M. S. Schmøkel, J. Overgaard and B. B. Iversen, *Acta Crystallogr. Sect. A*, 2012, **68**, 301–303.
81 J. Henn and K. Meindl, *Acta Crystallogr. Sect. A*, 2012, **68**, 304–304.
82 R. H. Blessing, *Acta Crystallogr. Sect. A*, 1995, **51**, 33–38.
83 D. Schwarzenbach, S. C. Abrahams, H. D. Flack, W. Gonschorek, T. Hahn, K. Huml, R. E. Marsh, E. Prince, B. E. Robertson, J. S. Rollett and A. J. C. Wilson, *Acta Crystallogr. Sect. A*, 1989, **45**, 63–75.
84 D. Watkin, *J. Appl. Crystallogr.*, 2008, **41**, 491–522.
85 P. Müller, *Crystallogr. Rev.*, 2009, **15**, 57–83.
86 B. Zarychta, J. Zaleski, J. Kyzioł, Z. Daszkiewicz and C. Jelsch, *Acta Crystallogr. Sect. B*, 2011, **67**, 250–262.
87 G. J. Kleywegt, *Acta Crystallogr. Sect. D*, 2000, **56**, 249–265.
88 P. Coppens(Project Reporter), J. Dam, S. Harkema, D. Feil, R. Feld, M. S. Lehmann, R. Goddard, C. Krüger, E. Hellner, H. Johansen, F. K. Larsen, T. F. Koetzle, R. K. McMullan, E. N. Maslen and E. D. Stevens, *Acta Crystallogr. Sect. A*, 1984, **40**, 184–195.
89 M. R. Probert, presented in part at the Electron Distribution & Chemical Bonding, Gordon Research Conference, Mount Holyoke College, MA, July, 2007.
90 T. Koritsanszky, A. Volkov and P. Coppens, *Acta Crystallogr. Sect. A*, 2002, **58**, 464–472.
91 Brock, C. P., Dunitz, J. D. and Hirshfeld, F. L, *Acta Crystallogr. Sect. B*, 1991, **47**, 789–797.
92 C. Jelsch, V. Pichon-Pesme, C. Lecomte and A. Aubry, *Acta Crystallogr. Sect. D*, 1998, **54**, 1306–1318.
93 C. Jelsch, T. Martha M., V. Lamzin, V. Pichon-Pesme, R. H. Blessing and C. Lecomte, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 3171–3176.
94 V. Pichon-Pesme, C. Lecomte and H. Lachekar, *J. Phys. Chem.*, 1995, **99**, 6242–6250.
95 B. Zarychta, V. Pichon-Pesme, B. Guillot, C. Lecomte and C. Jelsch, *Acta Crystallogr. Sect. A*, 2007, **63**, 108–125.
96 S. Domagala and C. Jelsch, *J. Appl. Crystallogr.*, 2008, **41**, 1140–1149.
97 P. M. Dominiak, A. Volkov, X. Li, M. Messerschmidt and P. Coppens, *J. Chem. Theory Comput.*, 2007, **3**, 232–247.
98 B. Dittrich, T. Koritsánszky and P. Luger, *Angew. Chem. Int. Ed.*, 2004, **43**, 2718–2721.
99 B. Dittrich, C. B. Hübschle, P. Luger and M. A. Spackman, *Acta Crystallogr. Sect. D*, 2006, **62**, 1325–1335.
100 J. M. Bąk, S. Domagała, C. Hübschle, C. Jelsch, B. Dittrich and P. M. Dominiak, *Acta Crystallogr. Sect. A*, 2011, **67**, 141–153.
101 B. Dittrich, C. B. Hübschle, J. J. Holstein and F. P. A. Fabbiani, *J. Appl. Crystallogr.*, 2009, **42**, 1110–1121.
102 W. F. Sanjuan-Szklarz, A. A. Hoser, M. Gutmann, A. Ø. Madsen and K. Woźniak, *IUCrJ*, 2016, **3**, 61–70.
103 V. Pichon-Pesme, C. Jelsch, B. Guillot and C. Lecomte, *Acta Crystallogr. Sect. A*, 2004, **60**, 204–208.
104 A. Volkov, T. Koritsanszky, X. Li and P. Coppens, *Acta Crystallogr. Sect. A*, 2004, **60**, 638–639.

105 S. K. J. Johnas, B. Dittrich, A. Meents, M. Messerschmidt and E. F. Weckert, *Acta Crystallogr. Sect. D*, 2009, **65**, 284–293.
106 E. Prince, *Mathematical Techniques in Crystallography and Materials Science*, Springer, Berlin, 2004, vol. 3.
107 G. M. Sheldrick, *Acta Crystallogr. Sect. C*, 2015, **71**, 3–8.
108 C. Jelsch, B. Guillot, A. Lagoutte and C. Lecomte, *J. Appl. Crystallogr.*, 2005, **38**, 38–54.
109 C. Jelsch, B. Guillot, S. Domagała, B. Fournier, I. Mata and C. Iordache, MOPRO 12.06 Documentation, http://crm2.univ-lorraine.fr/lab/fr/software/mopro/documentation-mopro/, (accessed 2 January 2018).
110 A. J. C. Wilson, *Acta Crystallogr. Sect. A*, 1976, **32**, 994–996.
111 B. Zarychta, J. Zaleski, J. Kyzioł, Z. Daszkiewicz and C. Jelsch, *Acta Crystallogr. B*, 2011, **67**, 250–262.
112 N. T. Johnson, H. Ott and M. R. Probert, *J. Appl. Crystallogr.*, 2018, **51**, 200–204.
113 T. E. Oliphant, *Comput. Sci. Eng.*, 2007, **9**, 10–20.
114 S. van der Walt, S. C. Colbert and G. Varoquaux, *Comput. Sci. Eng.*, 2011, **13**, 22–30.
115 J. D. Hunter, *Comput. Sci. Eng.*, 2007, **9**, 90–95.
116 M. Summerfield, *Rapid Gui Programming with Python and Qt: The Definitive Guide to Pyqt Programming*, Prentice Hall Press, Upper Saddle River, NJ, USA, First., 2007.
117 R. W. Grosse-Kunstleve, N. K. Sauter, N. W. Moriarty and P. D. Adams, *J. Appl. Crystallogr.*, 2002, **35**, 126–136.
118 Bruker AXS Inc., *XPREP*, Bruker AXS Inc., Madison, Wisconsin, USA, 2012.
119 Bruker AXS Inc., *XT*, Bruker AXS Inc., Madison, Wisconsin, USA, 2012.
120 Bruker AXS Inc., *XL*, Bruker AXS Inc., Madison, Wisconsin, USA, 2012.
121 P. Macchi, H.-B. Bürgi, A. S. Chimpri, J. Hauser and Z. Gál, *J. Appl. Crystallogr.*, 2011, **44**, 763–771.
122 A. B. Storm, C. Michaelsen, A. Oehr and C. Hoffmann, in *X-Ray Sources and Optics*, International Society for Optics and Photonics, 2004, vol. 5537, pp. 177–182.
123 J. D. Dunitz, *X-Ray Analysis and the Structure of Organic Molecules*, Cornell University Press, 2007.
124 Z. Chua, B. Zarychta, C. G. Gianopoulos, V. V. Zhurov and A. A. Pinkerton, *Acta Crystallogr. Sect. B*, 2017, **73**, 654–659.
125 J. M. Bąk, P. M. Dominiak, C. C. Wilson and K. Wozńiak, *Acta Crystallogr. Sect. A*, 2009, **65**, 490–500.
126 A. A. Hoser, P. M. Dominiak and K. Woźniak, *Acta Crystallogr. Sect. A*, 2009, **65**, 300–311.
127 V. Zavodnik, A. Stash, V. Tsirelson, R. de Vries and D. Feil, *Acta Crystallogr. Sect. B*, 1999, **55**, 45–54.
128 S. Swaminathan, B. M. Craven, M. A. Spackman and R. F. Stewart, *Acta Crystallogr. Sect. B*, 1984, **40**, 398–404.
129 A. Ø. Madsen, H. O. Sørensen, C. Flensburg, R. F. Stewart and S. Larsen, *Acta Crystallogr. Sect. A*, 2004, **60**, 550–561.
130 A. Ø. Madsen, *J. Appl. Crystallogr.*, 2006, **39**, 757–758.
131 P. Munshi, A. Ø. Madsen, M. A. Spackman, S. Larsen and R. Destro, *Acta Crystallogr. Sect. A*, 2008, **64**, 465–475.
132 W. Gao and A. Kahn, *Appl. Phys. Lett.*, 2001, **79**, 4040–4042.
133 P. Pingel, R. Schwarzl and D. Neher, *Appl. Phys. Lett.*, 2012, **100**, 143303.
134 J. E. Cochran, M. J. N. Junk, A. M. Glaudell, P. L. Miller, J. S. Cowart, M. F. Toney, C. J. Hawker, B. F. Chmelka and M. L. Chabinyc, *Macromolecules*, 2014, **47**, 6836–6846.
135 A. L. Sutton, B. F. Abrahams, D. M. D'Alessandro, T. A. Hudson, R. Robson and P. M. Usov, *CrystEngComm*, 2016, **18**, 8906–8914.
136 P. Hu, H. Li, Y. Li, H. Jiang and C. Kloc, *CrystEngComm*, 2017, **19**, 618–624.

137 H. Wolf, *Data Quality Bench-Marking for High Resolution Bragg Data*, (Ph.D), Georg-August-Universität Göttingen, March, 2015.
138 S. Mondal, S. J. Prathapa and S. van Smaalen, *Acta Crystallogr. Sect. A*, 2012, **68**, 568–581.
139 S. Jagannatha Prathapa, J. Held and S. van Smaalen, *Z. Anorg. Allg. Chem.*, 2013, **639**, 2047–2056.
140 R. F. W. Bader, P. J. MacDougall and C. D. H. Lau, *J. Am. Chem. Soc.*, 1984, **106**, 1594–1605.
141 F. L. Hirshfeld, *Theor. Chim. Acta*, 1977, **44**, 129–138.
142 J. C. Ma and D. A. Dougherty, *Chem. Rev.*, 1997, **97**, 1303–1324.
143 K. Shimizu, M. F. Costa Gomes, A. A. H. Pádua, L. P. N. Rebelo and J. N. Canongia Lopes, *J. Phys. Chem. B*, 2009, **113**, 9894–9900.
144 Rigaku Oxford Diffraction, *CrysAlis (Pro)*, Rigaku Oxford Diffraction, Yarnton, England, 2006.
145 G. Winter, *J. Appl. Crystallogr.*, 2010, **43**, 186–190.
146 A. Authier and G. Chapuis, *A Little Dictionary of Crystallography*, International Union of Crystallography, 2nd edn., 2017.
147 G. S. Nichol and W. Clegg, *CrystEngComm*, 2007, **9**, 959–960.
148 D. Singhal and W. Curatolo, *Adv. Drug Deliv. Rev.*, 2004, **56**, 335–347.
149 K. Raza, *SOJ Pharm. Pharm. Sci.*, 2014, **1**, 10.
150 M. Schmidtmann, L. J. Farrugia, D. S. Middlemiss, M. J. Gutmann, G. J. McIntyre and C. C. Wilson, *J. Phys. Chem. A*, 2009, **113**, 13985–13997.
151 M. A. Spackman and D. Jayatilaka, *CrystEngComm*, 2009, **11**, 19–32.
152 M. A. Spackman and P. G. Byrom, *Chem. Phys. Lett.*, 1997, **267**, 215–220.
153 M. J. Turner, J. J. McKinnon, S. K. Wolff, D. J. Grimwood, P. R. Spackman, D. Jayatilaka and M. A. Spackman, *CrystalExplorer17*, University of Western Australia, 2017.
154 P. Metrangolo and G. Resnati, *IUCrJ*, 2013, **1**, 5–7.
155 A. Mukherjee and G. R. Desiraju, *IUCrJ*, 2013, **1**, 49–60.
156 N. Ramasubbu, R. Parthasarathy and P. Murray-Rust, *J. Am. Chem. Soc.*, 1986, **108**, 4308–4314.
157 M. Barceló-Oliver, C. Estarellas, A. García-Raso, A. Terrón, A. Frontera, D. Quiñonero, I. Mata, E. Molins and P. M. Deyà, *CrystEngComm*, 2010, **12**, 3758–3767.
158 M. A. Spackman and J. J. McKinnon, *CrystEngComm*, 2002, **4**, 378–392.
159 D. J. Carter, P. Raiteri, K. R. Barnard, R. Gielink, M. Mocerino, B. W. Skelton, J. G. Vaughan, M. I. Ogden and A. L. Rohl, *CrystEngComm*, 2017, **19**, 2207–2215.
160 M. J. Turner, S. Grabowsky, D. Jayatilaka and M. A. Spackman, *J. Phys. Chem. Lett.*, 2014, **5**, 4249–4255.
161 C. F. Mackenzie, P. R. Spackman, D. Jayatilaka and M. A. Spackman, *IUCrJ*, 2017, **4**, 575–587.
162 J. Bernstein and J. M. Bernstein, *Polymorphism in Molecular Crystals*, Clarendon Press, 2002.
163 A. I. Isakov, E. N. Kotelnikova, L. Y. Kryuchkova and H. Lorenz, *Trans. Tianjin Univ.*, 2013, **19**, 86–91.
164 T. T.-D. Tran, P. H.-L. Tran, J.-B. Park and B.-J. Lee, *Arch. Pharm. Res.*, 2012, **35**, 1223–1230.
165 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. Sect. B*, 2016, **72**, 171–179.
166 Y. Krupskaya, M. Gibertini and A. F. Morpurgo, *Adv. Mater.*, 2015, **27**, 2453–2458.
167 T. Salzillo, M. Masino, G. Kociok-Köhn, D. Di Nuzzo, E. Venuti, R. G. Della Valle, D. Vanossi, C. Fontanesi, A. Girlando, A. Brillante and E. Da Como, *Cryst. Growth Des.*, 2016, **16**, 3028–3036.

168 T. J. Emge, M. Maxfield, D. O. Cowan and T. J. Kistenmacher, *Mol. Cryst. Liq. Cryst.*, 1981, **65**, 161–178.
169 M. I. Aroyo, *International Tables for Crystallography*, Wiley, 2nd edn., 2016, vol. A.
170 I. Y. Chernyshov, M. V. Vener, E. V. Feldman, D. Y. Paraschuk and A. Y. Sosorev, *J. Phys. Chem. Lett.*, 2017, **8**, 2875–2880.
171 L.-F. Ji, J.-X. Fan, S.-F. Zhang and A.-M. Ren, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3784–3794.
172 A. Y. Sosorev, *Phys. Chem. Chem. Phys.*, 2017, **19**, 25478–25486.
173 J. Bernstein, *J. Phys. Appl. Phys.*, 1993, **26**, B66.
174 G. R. Desiraju, *CrystEngComm*, 2007, **9**, 91–92.
175 K. M. Anderson and J. W. Steed, *CrystEngComm*, 2007, **9**, 328–330.
176 K. M. Steed and J. W. Steed, *Chem. Rev.*, 2015, **115**, 2895–2933.
177 K. M. Anderson, K. Afarinkia, H. Yu, A. E. Goeta and J. W. Steed, *Cryst. Growth Des.*, 2006, **6**, 2109–2113.
178 A. Lemmerer and C. Esterhuysen, *CrystEngComm*, 2011, **13**, 5773–5782.
179 C. P. Brock, *Acta Crystallogr. Sect. B*, 2016, **72**, 807–821.
180 A. Agarwal, A. Shaharyar, A. Kumar, M. S. Bhat and M. Mishra, *J. Clin. Orthop. Trauma*, 2015, **6**, 101–107.
181 V. Schomaker and K. N. Trueblood, *Acta Crystallogr. Sect. B*, 1968, **24**, 63–76.
182 J. Hvoslef, *Acta Crystallogr. Sect. B*, 1968, **24**, 23–35.
183 J. Hvoslef, *Acta Crystallogr. Sect. B*, 1968, **24**, 1431–1440.
184 J. R. Cheeseman, M. T. Carroll and R. F. W. Bader, *Chem. Phys. Lett.*, 1988, **143**, 450–458.
185 I. Love, *J. Phys. Chem. A*, 2006, **110**, 10507–10512.
186 R. E. Long, R. A. Sparks and K. N. Trueblood, *Acta Crystallogr.*, 1965, **18**, 932–939.
187 J. Ferraris, D. O. Cowan, V. Walatka and J. H. Perlstein, *J. Am. Chem. Soc.*, 1973, **95**, 948–949.
188 D. Jérome and H. J. Schulz, *Adv. Phys.*, 2002, **51**, 293–479.
189 D. S. Acker and W. R. Hertler, *J. Am. Chem. Soc.*, 1962, **84**, 3370–3374.
190 F. Iwasaki and A. Aihara, *Acta Crystallogr. Sect. B*, 1970, **26**, 91–97.
191 R. Kahn, R. Fourme, D. André and M. Renaud, *Acta Crystallogr. Sect. B*, 1973, **29**, 131–138.
192 B. P. Stoicheff, *Tetrahedron*, 1962, **17**, 135–145.
193 R. C. Wheland and E. L. Martin, *J. Org. Chem.*, 1975, **40**, 3101–3109.
194 D. R. Allan, H. Nowell, S. A. Barnett, M. R. Warren, A. Wilcox, J. Christensen, L. K. Saunders, A. Peach, M. T. Hooper, L. Zaja, S. Patel, L. Cahill, R. Marshall, S. Trimnell, A. J. Foster, T. Bates, S. Lay, M. A. Williams, P. V. Hathaway, G. Winter, M. Gerstel and R. W. Wooley, *Crystals*, 2017, **7**, 336.
195 N. T. Johnson, P. G. Waddell, W. Clegg and M. R. Probert, *Crystals*, 2017, **7**, 360.
196 DECTRIS, Pilatus CBF Header Specification, Version 1.4, https://www.dectris.com/technical_pilatus.html, (accessed 6 April 2015).
197 Bruker AXS Inc., *APEX2*, Bruker AXS Inc., Madison, Wisconsin, USA, 2007.
198 G. S. Almasi and A. Gottlieb, *Highly Parallel Computing*, Benjamin-Cummings Publishing Co., Inc., Redwood City, CA, USA, 1989.
199 D. G. Waterman, G. Winter, J. M. Parkhurst, L. Fuentes-Montero, J. Hattne, A. Brewster, N. K. Sauter and G. Evans, *CCP4 Newsl. Protein Crystallogr.*, 2013, *The DIALS framework for integration software*.
200 Bruker AXS Inc., *SAINT Software Reference Manual (Version 4)*, 1998.
201 H. J. Bernstein and A. P. Hammersley, in *International Tables for Crystallography Volume G: Definition and exchange of crystallographic data*, Springer, Dordrecht, 2006, pp. 37–43.

202 L. M. J. Kroon-Batenburg, presented in part at the Workshop on Metadata for raw data from X-ray diffraction and other structural techniques, Rovinj, Croatia, August, 2015.
203 J. M. Holton, S. Classen, K. A. Frankel and J. A. Tainer, *FEBS J.*, 2014, **281**, 4046–4060.
204 K. Diederichs, *Acta Crystallogr. Sect. D*, 2009, **65**, 535–542.
205 G. N. Phillips, *Biophys. J.*, 1995, **69**, 1281–1283.
206 J. W. Campbell, *CCP4 Newsl.*, 1996, *ROTGEN - A Program for Rotation Images Simulation and Analysis*.
207 A. R. Kolatkar, J. B. Clarage and G. N. Phillips, *Acta Crystallogr. Sect. D*, 1994, **50**, 210–218.
208 M. R. Probert, C. M. Robertson, J. A. Coome, J. A. K. Howard, B. C. Michell and A. E. Goeta, *J. Appl. Crystallogr.*, 2010, **43**, 1415–1418.
209 W. T. A. Harrison, J. Simpson and M. Weil, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2010, **66**, e1–e2.
210 T. Liu, Y.-X. Wang, Z.-W. Wang, Z.-P. Xie and J. Y. Zhu, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2010, **66**, e13–e14.
211 H. Zhong, S.-H. Duan, Y.-P. Hong, M.-L. Li, Y.-Q. Liu, C.-J. Luo, Q.-Y. Luo, S.-Z. Xiao, H.-L. Xie, Y.-P. Xu, X.-M. Yang, X.-R. Zeng and Q. Y. Zhong, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2009, **66**, e11–e12.
212 A. L. Spek, *Acta Crystallogr. Sect. D*, 2009, **65**, 148–155.
213 IUCr, Details of checkCIF/PLATON tests, http://journals.iucr.org/services/cif/datavalidation.html, (accessed 30 December 2017).
214 S. R. Hall, F. H. Allen and I. D. Brown, *Acta Crystallogr. Sect. A*, 1991, **47**, 655–685.
215 A. L. Spek, PLATON update history, http://www.cryst.chem.uu.nl/xraysoft/update_history_platon.html, (accessed 25 June 2018).
216 A. L. Spek, PLATON update history (Older), http://www.cryst.chem.uu.nl/xraysoft/old_update_history_platon.html, (accessed 26 March 2019).
217 IUCr, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2011, **67**, e16–e16.
218 IUCr, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2012, **68**, e10–e11.
219 IUCr, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2012, **68**, e14–e15.
220 IUCr, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2010, **66**, e21–e22.
221 IUCr, *Acta Crystallogr. Sect. E Struct. Rep. Online*, 2011, **67**, e14–e14.
222 Adam Marcus (amarcus41), Crystal myth: 11 more retractions from crystallography journal after 2010 fakery, http://retractionwatch.com/2011/02/28/crystal-myth-11-more-retractions-from-crystallography-journal-after-2010-fakery/#more-1878, (accessed 1 January 2018).
223 N. A. Vasilevsky, J. Minnier, M. A. Haendel and R. E. Champieux, *PeerJ*, 2017, **5**, e3208.
224 V. Stodden, J. Seiler and Z. Ma, *Proc. Natl. Acad. Sci.*, 2018, 201708290.
225 I. Shabalin, Z. Dauter, M. Jaskolski, W. Minor and A. Wlodawer, *Acta Crystallogr. Sect. D*, 2015, **71**, 1965–1979.
226 R. P. Joosten, T. Womack, G. Vriend and G. Bricogne, *Acta Crystallogr. Sect. D*, 2009, **65**, 176–185.
227 T. C. Terwilliger and G. Bricogne, *Acta Crystallogr. Sect. D*, 2014, **70**, 2533–2543.
228 D. Schwarzenbach, S. C. Abrahams, H. D. Flack, E. Prince and A. J. C. Wilson, *Acta Crystallogr. Sect. A*, 1995, **51**, 565–569.
229 J. M. Guss and B. McMahon, *Acta Crystallogr. Sect. D*, 2014, **70**, 2520–2532.
230 L. M. J. Kroon-Batenburg and J. R. Helliwell, *Acta Crystallogr. Sect. D*, 2014, **70**, 2502–2509.
231 S. W. M. Tanley, K. Diederichs, L. M. J. Kroon-Batenburg, A. M. M. Schreurs and J. R. Helliwell, *J. Synchrotron Radiat.*, 2013, **20**, 880–883.

232 J. R. Helliwell, B. McMahon, J. M. Guss and L. M. J. Kroon-Batenburg, *IUCrJ*, 2017, **4**, 714–722.
233 N. T. Johnson and M. R. Probert, presented in part at the Workshop on Metadata for raw data from X-ray diffraction and other structural techniques, Rovinj, Croatia, August, 2015.
234 M. Rossner and K. M. Yamada, *J. Cell Biol.*, 2004, **166**, 11–15.
235 J. Couzin-Frankel, *Science*, October, 2016, *Bringing image manipulation to light*.
236 M. Rossner, *The Scientist*, March, 2006, *How to guard against image fraud*.
237 B. Preneel, *Eur. Trans. Telecommun.*, 1994, **5**, 431–448.
238 A. L. Spek, presented in part at the American Crystallographic Association 66th Annual Meeting, Denver, CO, July, 2016.
239 A. L. Spek, Private Communication, 2015.
240 E. Rodrigo, B. G. Alcubilla, R. Sainz, J. L. G. Fierro, R. Ferritto and M. B. Cid, *Chem. Commun.*, 2014, **50**, 6270–6273.